FACULTY OF PSYCHOLOGY
AND EDUCATIONAL SCIENCES

Testing talent: validating tools for informed study choices in secondary education

Merel Dutry

Supervisor: Prof. Dr. Wouter Duyck
Co-supervisor: Dr. Nicolas Dirix & Prof. Dr. Eva Derous

A dissertation submitted to Ghent University in partial fulfilment of the requirements for the degree of Doctor of Psychology

Academic year 2024–2025

GHENT
UNIVERSITY

Table of Contents

Voorwoord

Je doctoraat afronden. Het is een prettige, en tegelijk ook een beetje vreemde gedachte. Vier jaar lang heb ik naar dit doel toegewerkt, en nu is dat doel bereikt.

Ik kan deze vier jaar best omschrijven als een intensieve reis, vol uitdagingen en ontdekkingen, met onverwachte wendingen, vermoeiende reisdagen, maar ook met adembenemende uitzichten en momenten van voldoening. Het was een reis die ik gelukkig niet alleen hoefde te maken, ik werd begeleid door ervaren gidsen, kreeg gezelschap van fantastische reisgenoten, en mijn supporters thuis zorgden voor dat comfortabele gevoel dat er een veilige haven is waar ik altijd terecht kan. Ik neem graag deze gelegenheid om hen allen te bedanken om mij tijdens deze reis te begeleiden en steunen.

Nicolas Dirix, mijn promotor, bedankt dat je als hoofdgids deze hele reis mee uitstippelde, mij de weg wees, begeleidde en steunde. Je constructieve feedback en toegankelijk leiderschap waren een houvast, en ik heb ontzettend veel van je geleerd. Dankzij jou bleef ik door het bos de bomen zien. Ook waardeer ik hoe je steeds luisterde naar mijn bezorgdheden en met heldere, logische argumenten rust en richting bracht.

Wouter Duyck, mijn (co-)promotor, bedankt dat je me vier jaar geleden de kans gaf om aan deze reis te beginnen. Je strategische inzichten en waardevolle feedback waren van onschatbare waarde. Bedankt voor je vertrouwen in mijn kunnen en voor het delen van jouw kennis en expertise.

Eva Derous, mijn co-promotor, bedankt voor je nauwe betrokkenheid gedurende mijn hele reis. In de eenzame beginfase tijdens corona hielp je me om niet te verdwalen op dit onbekende pad. Door me in contact te brengen met Maake en Annelies gaf je me reisgezellen in een anders stille en afgelegen startperiode. Dankjewel voor alle waardevolle feedback, voor je onvoorwaardelijke steun en aanmoediging om trouw te blijven aan mijn eigen koers.

Stijn Schelfout, mijn begeleider en onofficiële co-promotor, bedankt voor je inspirerende en doordachte feedback, die richting en houvast bood, ook wanneer het laat werd en de duisternis viel. Maar vooral bedankt voor je oprechte trots op ons hele team. Jouw enthousiasme over onze ideeën, onderzoeksresultaten en prestaties werkte aanstekelijk en motiveerde me om steeds het beste uit mezelf te halen.

Verder wil ik ook Bart Wille bedanken, lid van mijn begeleidingscommissie, voor zijn waardevolle feedback en de mogelijkheid om mijn doctoraat jaar na jaar voort te zetten. Daarnaast ook een oprechte dankjewel aan Klaas Bombeke, Jonas De Bruyne, Aleksandra Zheleva en Jelle Demanet, co-auteurs van Chapter 3. Bedankt voor jullie waardevolle inzichten en feedback, en voor het ter beschikking stellen van de Calibrain-applicatie.

Ook een oprechte dankjewel aan Mark Schittekatte, co-auteur van Chapter 1, en van meet af aan betrokken bij mijn reis en begaan met mij en het doctoraat. Bedankt voor het delen van jouw rijke kennis en expertise over intelligentie, en voor het vieren van de kleine mijlpalen onderweg. Ook een grote dank aan Dries Debeer, eveneens co-auteur van Chapter 1. Bedankt om jouw statistische kennis en expertise te delen, en voor jouw eeuwige geduld.

Heel graag bedank ik ook mijn lieve collega's en reisgezellen Mona, Sofie, Cathy en Mathias. Bedankt voor de gezellige tussenstops onderweg, voor jullie steun op momenten dat de weg zwaar voelde, voor de motivatie wanneer ik even dreigde te verdwalen, voor het delen van herkenbare doctoraatsavonturen, en voor alle fijne, eerlijke, en vaak hilarische gesprekken! Bedankt om deze reis zoveel aangenamer en lichter te maken.

Dankjewel Alexandra, voor je aanmoediging om een paper te schrijven over ons uitgebreide ICAR-onderzoek. Bedankt om mee aan die paper te schrijven, om sparringpartner te zijn en een team te vormen. Het was een samenwerking die voor mij precies op het juiste moment op mijn pad kwam. Ook bedankt voor de fijne gesprekken en gezellige lunches, die als rustpunten in de reis de nodige energie gaven om verder te gaan.

Ook bedankt aan de andere leden van het Duyck-Dirix lab, Samuël en Eleonore, voor de inspirerende labmeetings en waardevolle input. Bedankt ook aan Sam, Brent, Christophe V en Liesa voor de fijne lunches.

Dankjewel Christophe B, voor alle technische ondersteuning waardoor de reis soepel kon verlopen, en voor de gezellige koffiebreaks en lunchpauzes. En natuurlijk ook voor het verfrissende ijsje op die warme zomerdag. Bedankt ook aan Lies voor alle ondersteuning bij SAP en andere administratieve taken, en om ervoor te zorgen dat er altijd koffie was om de reis gaande te houden!

Bedankt Jens De Vos, Myriam Heeremans, Julie Beysens, Loena Nelissen, Jonas De Raeve, Valerie Hamal en nog vele andere collega's bij Voka, voor het oprichten van de leerstoel studieoriëntering en voor het uitbouwen van de Talentcenters. Bedankt om jullie schouders onder dit project te zetten en te streven naar een brede inbedding. Bedankt voor de korte lijn, de interessante vergaderingen en boeiende gesprekken. Bedankt voor de hele fijne samenwerking! Bedankt ook aan de onderwijsverstrekkers en de leden van de werkgroep voor jullie waardevolle input en leerrijke discussies. Ook bedankt aan alle leerlingen die drie jaar lang meewerkten aan het longitudinale onderzoek en aan de directies en leerkrachten die het onderzoek op hun school hebben gefaciliteerd en mee hebben opgevolgd.

Naast mijn gidsen, reisgezellen, en warme contacten onderweg wil ik ook een bijzondere dank uitspreken aan mijn supporters thuis. Bedankt Greet en Tom, mijn liefste en beste ouders! Bedankt om mij te steunen in elke stap van mijn leven en om zo fier te zijn op wie ik ben en wat ik doe. Bedankt om te geloven in mijn competenties zonder ooit enige druk te geven. Bedankt om mij zo graag te zien. Bedankt Milan, mijn liefste beste broeder. Bedankt om er altijd te zijn. Bedankt om altijd die lach op mijn gezicht te toveren. Bedankt Marie, de liefste schoonzus die ik me maar kon wensen. Bedankt Otto en Lou om met jullie kinderlijke verwondering en enthousiasme me telkens weer al mijn zorgen te doen vergeten. Bedankt ook aan mijn lieve schoonfamilie voor jullie oprechte interesse in mijn werk, en voor alle steun.

Bedankt aan mijn lieve vrienden, om mij te steunen en in mij te geloven. Om mij in het groot op TV af te spelen (waarvoor minder bedankt haha), en om mij de moed te geven om door te zetten in die laatste lastige weken. Maar vooral bedankt om mij zoveel vriendschap te geven en te zorgen voor al die onvergetelijke momenten samen over al die jaren heen! Merci. Bedankt ook aan Sammy, excel-tovenaar en consultant op zondagmorgen.

Tot slot wil ik Pieter bedanken, mijn partner en beste vriend, en ook de bedenker van de Ch-ICAR naam. Bedankt om er altijd voor mij te zijn. Om steeds opnieuw te luisteren naar hoe mijn dag was, ook al was die exact hetzelfde als de vorige. Om mijn reis van dichtbij te volgen en mij mee de weg te wijzen naar de eindbestemming. Bedankt om op tijd alles even in perspectief te plaatsen en om het volste vertrouwen te hebben in mijn capaciteiten. Bedankt om je liefste gekke zelf te zijn en het lichtpuntje in de soms lange lastige dagen.

Oprecht dikke merci allemaal om deze reis mogelijk te maken.

Gent, 18 februari 2025

Merel Dutry

Samenvatting

Ondanks het aangetoonde belang van loopbaanexploratie maken veel leerlingen in het secundair onderwijs belangrijke onderwijsbeslissingen met beperkte informatie en begeleiding. Deze situatie benadrukt de noodzaak van valide, contextspecifieke hulpmiddelen om leerlingen te ondersteunen bij complexe onderwijs- en loopbaankeuzes. Dit proefschrift speelt in op deze behoefte door innovatieve instrumenten te ontwikkelen, valideren en toe te passen die het studiekeuzeproces in het Vlaams secundair onderwijs verbeteren. Aan de hand van vier empirische hoofdstukken legt dit proefschrift de link tussen testvalidatie en praktische toepassingen en draagt het bij aan de wetenschappelijk onderbouwing van initiatieven zoals de Vlaamse Talentcenters. Door zich te richten op jonge adolescenten biedt dit proefschrift nieuwe inzichten in studieoriëntatie in het secundair onderwijs, een thema dat tot nu toe relatief weinig aandacht heeft gekregen.

In hoofdstuk 2 introduceerden we de Children's ICAR (Ch-ICAR), een aangepaste versie van de ICAR-test die is ontworpen om cognitieve vaardigheden bij kinderen te meten. Ons doel was om het ICAR-project uit te breiden, aangezien de geschiktheid van de ICAR-items voor kinderen nog niet eerder was aangetoond. Daarnaast wilden we inspelen op het gebrek aan psychometrisch robuuste, vrij toegankelijke en tijdsefficiënte cognitieve testen voor deze leeftijdsgroep. De resultaten ondersteunden de validiteit van de Ch-ICAR om de cognitieve vaardigheden van kinderen in kaart te brengen in een onderzoekscontext. De waargenomen correlatie met twee gevestigde intelligentietesten (de Raven’s 2 Progressive

Matrices en de CoVaT-CHC Basisversie), in combinatie met de gedocumenteerde relatie met schoolprestaties, leverde sterk bewijs voor de validiteit. Hoewel de totale Ch-ICAR sterke validiteit vertoonde, varieerden de afzonderlijke subtests in robuustheid, wat betekent dat we op dit moment voorzichtig moeten zijn bij het interpreteren van de individuele subtests.

In hoofdstuk 3 werd CalibrainVR-MR gevalideerd, een VR-test (*virtual reality*) voor het meten van mentale rotatievaardigheden bij leerlingen in het secundair onderwijs. Onze bevindingen toonden aan dat de test een betrouwbaar instrument is om de mentale rotatievaardigheden van leerlingen tussen elf en veertien jaar in kaart te brengen. De geobserveerde correlaties tussen de testresultaten en STEM-prestaties (*Science, Technology, Engineering, Mathematics*), vooral in STEM-studierichtingen, benadrukten de relevantie van de test nog verder. Opmerkelijk was dat er geen significant verschil werd gevonden in de mentale rotatiescores tussen leerlingen in STEM versus niet-STEM studierichtingen, wat mogelijkheden biedt om verborgen STEM-talent te identificeren. Daarnaast vertoonden zowel jongens als meisjes een vergelijkbare variantie in testscores, en werden er geen sekseverschillen waargenomen. Deze resultaten sluiten aan bij eerder onderzoek waaruit blijkt dat de traditioneel geobserveerde sekseverschillen in 2D-mentale rotatietaken afnemen in 3D-omgevingen. Samenvattend toonde hoofdstuk 3 het potentieel van CalibrainVR-MR en vergelijkbare VR-gebaseerde tools om de beoordeling van ruimtelijke vaardigheden te verbeteren. Deze tools maken de beoordeling niet alleen realistischer en eerlijker, maar ze bieden ook een nieuwe manier om STEM-talent te identificeren.

In hoofdstuk 4 onderzochten we de persoons-omgeving (PE) interesse fit in het secundair onderwijs met behulp van de SIMON-jr. Onze resultaten bevestigden de betrouwbaarheid en validiteit van de SIMON-jr en benadrukten dat het een degelijk en fair instrument is om leerlingen te ondersteunen bij hun studiekeuze. We observeerden dat middelbare scholieren over het algemeen studierichtingen kiezen die aansluiten bij hun interesses, waarbij meisjes een sterkere interesse fit vertoonden dan jongens. De gemiddelde interesse fit was echter lager dan in het hoger onderwijs, wat aansluit bij de bevinding dat interesses in de vroege adolescentie minder gedifferentieerd zijn en specifieker worden met de leeftijd. Daarnaast bleek er in deze fase geen verband te zijn tussen interesse fit en academische prestaties. Deze nulrelatie kan waarschijnlijk deels verklaard worden door de brede, algemene aard van de studierichtingen in het secundair onderwijs, die bedoeld zijn om leerlingen voor te bereiden op uiteenlopende loopbanen. Door dit gebrek aan specialisatie ontstaat er meer variatie in interesseprofielen binnen studierichtingen, wat leidt tot zwakkere verbanden met academische prestaties.

In hoofdstuk 5 onderzochten we hoe de PE interesse fit tijdens de vroege adolescentie ontwikkelt binnen het Vlaamse onderwijssysteem. Rekening houdend met het feit dat interesses in deze fase nog in ontwikkeling zijn, gebruikten we een longitudinaal onderzoeksdesign om te bestuderen hoe interesses en de fit met studierichtingen in de loop van de tijd evolueren. Onze resultaten toonden aan dat de interesse fit met de gekozen studierichting stabiel bleef, wat opvallend is gezien de veranderingen in afzonderlijke interessedimensies. Daarentegen nam de fit met niet-gekozen studierichtingen af, al was deze afname alleen bij jongens significant. Voor zowel jongens als meisjes werd het verschil tussen de fit met gekozen en niet-gekozen studierichtingen steeds groter, voornamelijk door de dalende fit met niet-gekozen opties. Deze bevindingen benadrukken de complexiteit van interesseontwikkeling en de wisselwerking tussen individuele en structurele factoren bij studiekeuzes. SIMON-jr bleek een waardevol instrument om deze dynamieken in kaart te brengen, wat zijn nut bij de begeleiding van jongeren in hun studiekeuzeproces ondersteunt.

Concluderend biedt dit proefschrift waardevolle theoretische en praktische inzichten in studieoriëntatie en loopbaanverkenningsprocessen in het secundair onderwijs. In dit proefschrift valideerden we innovatieve instrumenten, zoals de Ch-ICAR, CalibrainVR-MR, en SIMON-jr, die betrouwbare en eerlijke metingen opleveren van vaardigheden en interesses die een rol spelen bij studiekeuzes en schoolprestaties. Daarnaast benadrukken onze bevindingen het potentieel van exploratietools zoals het Vlaamse Talentcenter, dat een gestructureerd kader biedt om leerlingen te ondersteunen bij het maken van weloverwogen studiekeuzes.

Summary

Despite the recognized importance of career exploration, many secondary education students face the challenge of making early educational decisions with limited information and guidance. This tendency highlights the need for valid, context-specific tools to support students in navigating the complexities of educational and career choices. This dissertation addressed this need by developing, validating, and applying innovative tools designed to improve the study choice process in Flemish secondary education. Through four empirical chapters, this dissertation bridged test validation with real-world applications, particularly within initiatives like the Flemish Talentcenters. By focusing on early adolescents, this dissertation offered novel insights into study orientation in secondary education—an area that has received relatively little attention.

In Chapter 2, we introduced the Children's ICAR (Ch-ICAR), an adaptation of the ICAR project designed to assess cognitive abilities in children. Our goal was to expand the ICAR's scope, as its suitability for children had not been previously evaluated. We also aimed to address the lack of psychometrically robust, freely accessible, and time-efficient cognitive assessments for this age group. Our results confirmed the effectiveness of the full Ch-ICAR as a research tool for measuring children's cognitive abilities. The correlation with two well-established intelligence tests (i.e., the Raven's 2 Progressive Matrices and the CoVaT-CHC Basic Version), along with the established relationship with academic performance, provided

evidence of its validity. While the Ch-ICAR demonstrated overall validity, individual subtests varied in robustness, prompting caution when interpreting subtests independently at this stage.

Chapter 3 validated CalibrainVR-MR, a virtual reality (VR)-based tool for assessing mental rotation abilities in secondary education students. Our findings confirmed its utility as a reliable measurement of mental rotation abilities in students aged eleven to fourteen. The test's moderate correlations with STEM performance, particularly in STEM programs, further underscored its relevance. Interestingly, no significant difference in mental rotation scores was found between STEM and non-STEM students, suggesting untapped potential for identifying future STEM talent. Additionally, both boys and girls showed equal variance in scores, with no sex differences observed, consistent with research indicating that sex differences in 2D mental rotation tasks tend to diminish in 3D environments. Overall, Chapter 3 underscored the potential of CalibrainVR-MR and similar VR-based tools in advancing the assessment of spatial abilities. These tools not only improve the realism and fairness of evaluations but also offer a novel way to identify STEM talent that may be overlooked by traditional assessments.

In Chapter 4 we examined person-environment (PE) interest fit in secondary education using the SIMON-jr instrument. Our results confirmed the reliability and validity of the SIMON-jr, establishing it as a robust and equitable tool for guiding students in their educational choices. We found that secondary students generally select education programs that fit their interests, with girls showing a stronger interest fit than boys. However, the mean interest fit was lower than that typically observed in higher education, supporting the idea that vocational interests are less differentiated in early adolescence and become more specific with age. Additionally, interest fit did not correlate with academic achievement at this stage. The broader, more general nature of secondary education programs likely contributes to this absence, as these programs are designed to prepare students for diverse career paths. This lack of specialization results in greater variation in RIASEC profiles within programs, leading to weaker associations with academic performance.

In Chapter 5, we explored the longitudinal development of PE interest fit during early adolescence within the Flemish educational system. Acknowledging the developmental nature of vocational interests during this stage, we implemented a longitudinal study design to examine how interests and their fit with educational choices evolve over time. Our findings revealed that PE fit with the chosen program remained stable over time, which is notable given the changes observed in singular interest dimensions. In contrast, PE fit with unchosen programs decreased over time, although only at a significant level for boys. For both boys and girls, the difference between the fit with chosen and unchosen programs became increasingly pronounced, primarily driven by the decline in fit with unchosen options. These findings highlight the complexity of vocational interest development and the interplay of individual and structural factors. The SIMON-jr instrument proved valuable in capturing these dynamics, reinforcing its utility in guiding early adolescents' educational choices.

In conclusion, this dissertation provided important theoretical and empirical insights into study orientation and career exploration processes in secondary education. Across four empirical studies, our research validated innovative tools like the Ch-ICAR, CalibrainVR-MR, and SIMON-jr. These tools enable the reliable and developmentally appropriate assessment of abilities and interests linked to educational decisions and academic performance, while also promoting inclusion. Finally, our findings highlight the potential of exploration tools like the Flemish Talentcenter, ultimately supporting students in making informed study choices.

Chapter 1.

Introduction

Selecting a secondary education program is a crucial milestone in a student's academic and personal development (Seghers et al., 2019). To make optimal career decisions, career choice theorists emphasize the importance of exploring both the self and the environment (e.g. Gottfredson, 1981; Holland, 1997). Yet, in practice, many students face this decision with limited information. For instance, 12-year-olds often lack a clear understanding of what specific fields of study entail or fail to grasp the long-term implications of their choices (Goosen et al., 2017). Social inequalities further complicate the decision-making process, as knowledge of the education system and access to critical information are unevenly distributed across social classes (Ball, 2003; Draelants, 2014; Seghers et al., 2019). This social background also influences teacher recommendations, as teachers are more likely to steer students from advantaged families toward academic tracks, even when their performance is comparable to that of their less privileged peers (Seghers et al., 2019). Additionally, decisions about secondary education are frequently driven by perceived cognitive ability and academic demands, with insufficient attention to students' domain-specific interests. This misalignment is concerning, as a strong congruence between personal interests and the educational environment is associated with higher rates of study completion, career satisfaction, and overall success (de Vries et al., 2024; Nye et al., 2012).

Beyond individual and social barriers, systemic challenges within the education system amplify the need for enhanced study-orientation processes. A significant challenge is the *waterfall effect*, where students initially enroll in more prestigious academic tracks but later switch to more practical programs after struggling to meet the demands of their initial choices (Van Praag et al., 2015). These transfers are often perceived as failures, contributing to higher dropout rates and diminished academic outcomes. A related challenge is the growing demand for skilled technical workers, which contrasts with the historical undervaluation of vocational education (Vandenbroeck et al., 2024). This labor market mismatch, where students' educational paths are poorly aligned with labor market demands, poses challenges to sustainable economic growth.

These challenges underscore the urgent need for valid, context-specific tools to assist secondary education students in making well-informed educational decisions that align with their interests, talents, and aspirations. This dissertation addresses this need by focusing on the development, validation, and application of tools designed to improve the study choice process in secondary education. These tools are implemented within the Flemish Talentcenters (https://talentcenter.vlaanderen/), an initiative by Voka vzw, the Flemish Network of Enterprises and one of the largest business organizations in Flanders, Belgium.

To provide a clearer understanding of the research context, we begin this introduction by offering a brief overview of the Flemish education system, highlighting the specific dynamics and challenges within this regional framework.

Educational system in Flanders

In Flanders, compulsory education begins at age five and typically comprises one year of kindergarten, followed by six years of primary and six years of secondary education (Flemish Government, n.d.-1). While kindergarten and primary education follow a unified curriculum, secondary education is structured into three, two-year cycles with increasing specialization. Upon entering secondary school (ages 12-13), students enroll in either the A-stream or the B-stream. The B-stream is designed for students who did not met the minimum requirements of primary education. Most students, however, enroll in the A-stream, with 86% of first-year secondary students following this path during the 2023-2024 academic year (Department of Education and Training, 2023). While the A-stream offers a standardized curriculum aimed at preparing students for later educational choices, early tracking occurs through elective courses. Many schools provide these electives in line with the specific study programs they offer in the second cycle, effectively beginning the specialization process during the transition from primary to secondary school (Seghers et al., 2019).

At the start of the second cycle (ages 14-15), students must choose a specific study program from one of eight domains: Language and Culture, STEM, Art and Creation, Agriculture and Horticulture, Economy and Organization, Society and Well-being, Sport, or Food and Catering. Additionally, students select one of three finalities: academic (preparation for higher education), vocational (preparation for workforce entry), or dual (preparation for both). Within the vocational finality, students have the option to enroll in part-time vocational education, which blends classroom instruction with hands-on workplace experience. Individual study programs are categorized within a domain-by-finality framework based on content similarities. By the start of the third cycle (ages 16-17), students are required to further specialize their field of study. Attendance in secondary school is mandatory until the age of eighteen, and students who leave school before completing their six years of study and obtaining a diploma or qualification are considered to have dropped out.

Each year concludes with the awarding of a certificate that determines whether a student may continue their current educational trajectory. A Certificate A grants students unrestricted progression, while a Certificate B allows advancement to the next grade but limits access to certain fields of study. A certificate C indicates that the student must repeat the year (Flemish Government, n.d.-2). These decisions are based on Grade Point Average (GPA), which is derived from tests and assignments devised by individual teachers. Notably, there are no centralized or standardized assessments, such as national exams, to guide these evaluations[1] (Stevens, 2007; Van Houtte & Demanet, 2016).

[1]Even though *Vlaamse toetsen* have recently been introduced, these tests are not used for individual assessments. (Flemish government, n.d.-3).

The Importance of Career Exploration

Every year, approximately 18,000 secondary students in Flanders change their field of study (De Windt, 2020). This high rate of school mobility highlights the challenges many students face in finding and committing to a suitable educational pathway. These challenges are further underscored by concerning statistics on school dropout. In the 2021-2022 academic year, 14% of students in Flanders left secondary education without qualification (Statistics Flanders, 2024). Notably, this percentage is rising, particularly in urban areas (Voka, 2024). Research indicates a strong overlap between factors contributing to school mobility and those associated with dropout (Degroote et al., 2020). Both issues stem from common underlying causes, including low levels of social and academic engagement in school (Fall & Roberts, 2012; Janosz et al., 2008; Rumberger & Rotermund, 2012). Addressing these root causes of disengagement and demotivation is essential, as the personal and societal consequences of school dropout are far-reaching (Anttila et al., 2022). On an individual level, leaving school prematurely imposes significant barriers to later life adjustment. Dropouts are more likely to face unemployment and have limited career prospects compared to individuals with higher qualifications (OECD, 2020). Additionally, dropping out negatively impacts health and well-being, compounding the challenges faced by these individuals (Griep et al., 2016; Ramsdal et al., 2018). There is also a significant impact for society, as dropout leads to a loss of human capital (Lee & Burkam, 2003) and undermines policymakers' efforts to achieve sustainable economic growth (De Witte et al., 2013). Moreover, high rates of student mobility and dropout also affect the non-mobile student population, as frequent student turnover disrupts the classroom environment, making it challenging for teachers to introduce new topics and maintain a consistent learning pace (Degroote et al., 2020). This hinders the academic progress of all students (Van Praag et al., 2015).

In addition, the *waterfall system*, a persistent phenomenon within the Flemish education system (Van Praag et al., 2015), exacerbates disengagement and school mobility. This system refers to the tendency of students to initially enroll in the academic finality, regarded as the most prestigious and demanding track, before transferring to more practical, vocational tracks after struggling to meet the demands of their initial choice (Goosen et al., 2017). This transfer pattern reflects a hierarchical structure in which vocational education is often seen as a fallback rather than a valid, positive choice (Spruyt, 2014). Such a hierarchy undermines students' engagement and self-esteem, as transfers are often perceived as failures rather than opportunities to align educational paths with their interests and strengths. This perception further diminishes motivation and contributes to dropout rates.

Notably, the growing demand for skilled technical workers contrasts with the undervaluation of vocational education (Vandenbroeck et al., 2024). Encouragingly, this perception is shifting, as nearly 59% of secondary students in Flanders now pursue technical or vocational tracks, marking a six-year trend of rising enrollments (Verhaeghe & Lamote, 2025). This shift seems to reflect reforms since 2019-2020 that redefined study fields into clear finalities—academic, vocational, and dual—helping dispel stigmas and position all pathways as equally valuable (De Witte, KU Leuven). With secondary education becoming more practice-oriented, students are better prepared for the labor market without losing access to higher education, particularly in dual-finality programs. While these trends are promising, continued efforts to promote the value of technical and vocational education are vital.

Supporting students in making informed, positive study choices plays a crucial role in sustaining this momentum. Career exploration, as a proactive and adaptive process, equips individuals with the tools to align their personal interests and abilities with the opportunities available in the world of work. By fostering early engagement in career exploration, students can make better-aligned academic and career decisions, ultimately leading to more satisfying career outcomes (Hirschi, 2009; Lent & Brown, 2013; Savickas, 2013). This approach also has the potential to address systemic challenges, such as high dropout rates and the persistent waterfall system, by guiding students toward pathways that resonate with their interests and strengths.

Career construction theory offers a valuable framework for understanding the importance of career exploration. The theory emphasizes that individuals who engage in continuous self-discovery and adaptation are more likely to achieve a meaningful alignment between their personal characteristics and the demands of their environments (Jiang et al., 2019). This iterative process is particularly vital during adolescence, as individuals begin to formulate career goals, acquire occupational knowledge, and develop their vocational identity (Savickas, 2002, 2013). During these formative years, structured career exploration provides the foundation for long-term vocational development, equipping students with the tools to navigate complex decisions and transitions. As individuals enter early adulthood, career exploration continues to refine their understanding of career identity, enabling them to make informed choices during the critical transition from education to the workforce. Even in later stages of adulthood, this process remains relevant, helping individuals adapt to evolving aspirations and labor market demands (Neureiter & Traut-Mattausch, 2017).

In light of these insights, the need for structured career exploration becomes evident, particularly in the context of secondary education. The challenges faced by students in secondary education underscore the importance of providing context-specific tools that foster informed and proactive career exploration. While numerous commendable initiatives already exist to support study exploration and career orientation in secondary education in Flanders, there remains a need for an accessible, scientifically grounded tool that complements these existing efforts. The Flemish Talentcenter aims to address this gap by supporting students aged eleven to fourteen in identifying their interests, strengths, and potential educational paths through a range of scientifically validated assessments and activities. The guidance provided by the Talentcenter supports both early decisions, such as selecting a school and elective courses in the first two years of secondary education, and the choice of a study program in the third year. Designed as a supplementary resource to support the study choice process in lower secondary education, the Talentcenter aims to enhance existing counseling practices rather than replacing them. Importantly, the Talentcenter does not seek to dictate students' future paths or infringe on the autonomy of parents and students in making educational decisions. Instead, it offers a comprehensive personalized overview of possibilities, helping to align educational options with students' preferences and strengths.

Development of the Flemish Talentcenters

The development of the Flemish Talentcenters started with the careful selection of constructs considered essential for ensuring study choice satisfaction and academic success. This selection process was guided by foundational career choice theories, which emphasize the importance of exploring both the self and the environment to make well-informed decisions (e.g., Gottfredson, 1981; Holland, 1997). Central to these theories is the principle of person-environment fit (Dawis, 2005; Schneider, 1987), which posits that students who make realistic choices based on their unique characteristics and the demands of their educational environment are more likely to succeed academically and professionally.

Interests (Hanna et al., 2021; Holland, 1997; Schelfhout et al., 2021) and competencies, including skills and abilities (Roth et al., 2015; Tikhomirova et al., 2020), have been identified as essential attributes for aligning individuals with suitable educational and career paths. Additionally, recent studies emphasize the significant role of non-cognitive factors, such as motivation, self-efficacy, and persistence, in shaping career decisions (e.g., Fonteyne et al., 2017; Credé & Kuncel, 2008; Lipnevich & Roberts, 2012). To comprehensively address these factors, the Talentcenter test battery incorporates both cognitive and non-cognitive constructs. Considering the young age of the target population (11–14 years), the test battery was carefully designed to balance comprehensiveness with manageability. The final test battery includes measures of vocational interests, mathematical and reasoning abilities, vocabulary, fine motor skills, spatial ability, technical insight, motivation, self-efficacy, and persistence.

The next step involved identifying suitable instruments to measure these constructs, guided by three criteria: (1) appropriateness for the target age group, (2) demonstrated validity and reliability, and (3) feasibility for long-term implementation, which emphasized freely available tools. Existing tests meeting these criteria were incorporated, with some accessed through partnerships. For three constructs—reasoning ability, vocational interests, and spatial ability—no suitable tests meeting the established criteria were available at the time of inception. Consequently, custom tests were developed to address the specific needs of the Talentcenter. These tests were piloted and refined before being incorporated into the longitudinal research. The development and validation processes for these measures are described in detail in Chapters 2 to 4. By developing these tests, we addressed the practical needs of the Talentcenter while simultaneously filling two significant gaps in the academic literature. First, much of the existing research on career exploration has focused on higher education or adulthood (Su, 2020), leaving early adolescents largely understudied. Our work extends existing literature by focusing on this younger age group and provides valuable insights for study counselors, helping them improve decision-making guidance and develop career guidance tools tailored to the unique needs of early adolescents. Second, our research responds to calls for better data and accessible resources to advance the study of key theoretical constructs (Uttal et al., 2024). Open-access instruments, like those developed here, enable researchers to conduct studies, share findings, and collaboratively advance the field (Condon & Revelle, 2014). An overview of all instruments used in the Talentcenter is presented in Table 1.

Table 1

Used Instruments in the Talentcenters

Construct	**Instrument**	**Cronbach's alpha**
Motivation	Adapted version of the Academic Self-Regulation Scale (Ryan & Connell, 1989) by Vansteenkiste and colleagues (2005)	Identified regulation: α = .80 (*N* items = 4) Intrinsic motivation: α = .84 (*N* items = 4)
Self-efficacy	Academic self-efficacy scale (Vantieghem & Van Houtte, 2015)	α = .81 (*N* items = 8)
Persistence	Persistence-items from the Children's Perceived Use of Self-Regulated Learning Inventory (CP-SRLI) (Vandevelde et al., 2013)	α = .84 (*N* items = 6)
Technical insight	Items from mijnstemcheck.be, developed by Constructive, Woodwize, Volta, Cego, KU Leuven & Vlaanderen	α = .63 (*N* items = 5)
Mathematical ability	Shortened version of the *Cognitive Developmental Skills in Arithmetics* (*Cognitieve Deelhandelingen van het Rekenen* [CDR]; Desoete & Roeyers, 2006)	α = .72 (*N* items = 18)
Vocabulary	Revised version of the Dutch online youth-word test (Brysbaert et al., 2014) as used in Bourdeaud'hui et al. (2020)	α = .70 (*N* items = 45)

Fine motor skills	Motor performance series (MLS) of Schoppe (1974), in the version of Schuhfried (2018)	See Aschenbrenner et al. (2012)[a]
Reasoning ability	Ch-ICAR (Dutry et al., 2025)	α = .82 (*N* items = 27)
Spatial ability	CalibrainVR-MR, adapted version of CaliBrainVR (Zheleva et al., 2023) by imect-MICT-UGent & Howest	α = .64 (*N* items = 34)
Vocational interests	SIMON-jr (Dutry et al., *In Preparation*)	*N* items per interest dimension = 6. α in order of RIASEC acronym = .80, .81, .85, .77, .70, .65.

Note. The Cronbach's alpha values are based on data from the Talentcenter participants collected between April 2023 and December 2024 (N = 8,646).
[a] For motor skills, test-retest reliability (along with inter-rater reliability) is the most commonly used reliability measure. Since test-retest reliability data are not available for our sample, we refer to the test-retest reliability measures provided in the Schuhfried manual (Aschenbrenner et al., 2012).

Data Collection

The next step in the development of the Talentcenters involved a longitudinal study with two primary objectives: (1) validating the selected instruments within the target population and (2) examining the relationship between test results, academic achievement, and satisfaction with the study choice made in the third year of secondary education. The research protocol[2] was approved by the Ethical Commission of the Faculty of Psychology and Educational Sciences at Ghent University (reference number 2021/59).

While the ideal starting point for the longitudinal study would have been the final year of primary education (as these students are part of the Talentcenter's target population), practical and methodological constraints made this approach less feasible. For instance,

[2]Entitled: "Study orientation in secondary education".

response rates would likely decline significantly between waves due to students dispersing to different secondary schools. Additionally, it was essential to ensure a representative sample of secondary schools to capture the diversity of educational contexts and strengthen the generalizability of the findings. Starting in primary education would have offered much less control over the diversity of the sample. By instead tracking first-year secondary students from the beginning of the academic year, the study's findings could still be extrapolated to primary school students while maintaining a more representative sample.

The first wave of data collection took place in the fall of the 2021–2022 academic year and involved 899 first-year secondary students who received consent from their parents to participate. The second wave occurred in the spring of 2023, involving the same cohort in their second year of secondary education. By the third and final wave, conducted in January-February of the 2023–2024 academic year, 545 students completed the assessments. Scheduling this final wave after the Christmas break ensured that students had completed at least one exam period and could provide meaningful feedback on their study choice satisfaction.

At the end of each school year, academic achievement data were also collected from school management to complement the survey responses[3]. This structured approach ensured that the longitudinal study captured comprehensive data on the connections between test results, academic performance, and study choice satisfaction.

Further development of the Talentcenters

In April 2023—prior to the final wave of data collection—we established a pilot Talentcenter at Technopolis, Mechelen. This pilot served to evaluate the concept's appeal and refine its operationalization. Furthermore, this pilot Talentcenter provided a new stream of

[3]In the third year of secondary education, semester 1 results were requested instead of waiting until the end of the school year.

data, which enabled further investigation into the validity and reliability of our tests. This new cohort also provided a fresh sample for demonstrating the predictive validity of CalibrainVR-MR (see Chapter 3).

The initial personalized feedback report offered valuable insights into students' test results. Indeed, the validity of the individual tests within our target population had already been confirmed through the ongoing longitudinal research, which also allowed us to establish reference categories using the available data. Additionally, we had conducted a cross-sectional study to map the interest profiles of study programs available in the third year of secondary education. The process for determining the interest profiles of these study environments is detailed in Chapter 4. These profiles enabled us to calculate the degree of fit between a student's interests and the characteristics of various study programs, allowing us to suggest study programs that match individual interest profiles.

Once the final data collection was completed, we finalized the Talentcenter feedback model, enabling us to provide suggestions for study programs aligned with students' interests as well as performance on all instruments. A new feedback model was implemented in the Talentcenters at the start of the academic year 2024-2025 and is based on the data collected during the longitudinal study. Although the overarching feedback model is valuable for the exploration process and scientifically intriguing, it is not included in this dissertation. This work focuses specifically on the initial validation of the newly developed components that underpin the feedback model. The other tests utilized in the Talentcenter feedback model had already been validated within the same population and demonstrated adequate reliability in the current context. Internal consistency values for each test are provided in Table 1. These additional tests are not further discussed in this dissertation.

Overview of the Chapters

This dissertation documents the development, validation, and application of innovative instruments aimed at enhancing the study choice process in secondary education. Grounded in the academic pursuit of refining this critical domain, the research presented here contributes novel methodologies and tools tailored to the needs of the Flemish Talentcenters. These efforts are rooted in career exploration theory (Jiang et al., 2019), which emphasizes the interplay between personal and contextual factors in shaping individuals' career-related behaviors (e.g., Flum & Blustein, 2000; Lent & Brown, 2013; Savickas, 2002). By fostering self- and environmental exploration, the instruments developed in this dissertation aim to promote personal agency in career self-management, a central tenet of the Career Self-Management (CSM) framework (Lent & Brown, 2013).

Chapter 2 focuses on the validation of the Children's ICAR (Ch-ICAR), a public domain measure of cognitive ability for children aged 11–14. Cognitive ability, defined as the capacity to acquire, process, organize, and apply knowledge effectively (Gottfredson, 1997), is a key predictor of outcomes such as academic achievement (Roth et al., 2015; Tikhomirova et al., 2020) and work performance (Sackett et al., 2017; Schmidt, 2016). Cognitive ability also serves as an important control variable across various research domains. However, widely used cognitive ability tests often have practical limitations, such as lengthy administration times and high costs, making them less suitable for many research settings. To address these challenges, the International Cognitive Ability Resource (ICAR; Revelle et al., 2020) was developed as a public domain, online measure of cognitive ability. The ICAR provides psychometrically robust subtests (Revelle et al., 2017) with diverse difficulty levels. Notably, the suitability of ICAR for children was not yet examined. Given the rapid cognitive development during childhood and adolescence, it is essential to acknowledge that ICAR items originally designed for adults may not effectively assess the cognitive abilities of children (Flanagan & McDonough, 2018). Therefore, we aimed to create a child-appropriate version of the ICAR, the Ch-ICAR, combining the practical benefits of the ICAR with the specific requirements for measuring cognitive ability in children aged 11–14. Additionally, given that the ICAR was not explicitly based on a theoretical framework like the Cattell-Horn-Carroll (CHC) model of intelligence (Schneider & McGrew, 2012), we examined which broad cognitive abilities are captured by its subtests. Chapter 2 begins by describing the compilation of the Ch-ICAR, based on a pilot study involving 99 primary school pupils. This section is followed by Study 1, which examines the psychometric properties of the Ch-ICAR in a sample of 820 secondary school pupils. Finally, Study 2 addresses the construct validity of the Ch-ICAR by cross-validating it against two established measures: Raven's 2 Progressive Matrices and the Flemish CoVaT-CHC Basic Version. These analyses were conducted using samples of 91 secondary school pupils and 96 primary school pupils, respectively.

Chapter 3 introduces CalibrainVR-MR, an innovative virtual-reality-based test for measuring mental rotation abilities in adolescents. Mental rotation refers to the capacity to mentally manipulate 2D and 3D objects (Linn & Petersen, 1985), a skill classified within the domain of spatial abilities—a recognized factor of human intelligence (Hegarty & Waller, 2005). Traditionally, mental rotation ability has been assessed using 2D media-based tests. However, these assessments often face challenges, including ambiguities in 3D technical drawings and issues with occlusion (Bartlett & Camba, 2023). Recent advancements in VR technology provide exciting opportunities to overcome these challenges and enhance the assessment of mental rotation abilities (Bartlett et al., 2024). However, while VR and augmented reality (AR) technologies have been widely used to train spatial skills across various age groups (Piri & Cagiltay, 2024), their integration into formal assessments remains underdeveloped (Bartlett et al., 2024). Additionally, most validation studies of VR-based mental rotation tests (Bartlett et al., 2024; Lochhead et al., 2022) have focused on adult populations, leaving a significant gap in research involving younger students. To address this gap, we validated CalibrainVR-MR using a large sample of 1,223 students aged 11 to 14. Additionally, we analyzed longitudinal academic performance data from a subset of 131 students to explore the relationship between CalibrainVR-MR results and academic outcomes in both STEM and non-STEM fields. This analysis is grounded in the well-documented role of mental rotation as a key predictor of success in STEM disciplines (Cheng & Mix, 2014; Sorby et al., 2013; Veurink & Sorby, 2019), as well as in vocational education (Lee, 2020). By evaluating the test's potential as a domain-specific predictive tool, this study provides valuable insights into its practical applications in educational contexts.

Chapter 4 presents the validation of SIMON-jr, a newly developed, publicly available tool in Dutch, designed to assess vocational interests in students aged 11–14. This chapter also introduces an initial application of SIMON-jr, examining how students' vocational interests in the first year of secondary education (ages 12–13) align with their chosen study programs by the third year (ages 14–15), and how this Person-Environment (PE) interest fit (Schelfhout et al., 2021) is related to academic achievement. Additionally, we investigate potential gender differences in interest fit, offering nuanced insights into interest-based decision-making. While vocational interests and PE interest fit are well-established factors influencing academic and career success, much of the existing literature has concentrated on adults or university students. This focus leaves a notable gap in understanding how these factors apply to younger populations, particularly early adolescents (Seghers et al., 2019). This gap is significant because PE interest fit has proven valuable in career education for older individuals, fostering self-reflection and exploration (Hanna & Rounds, 2020; Schelfhout et al., 2021). Therefore, it is crucial to extend this investigation into the context of secondary education, where the potential to guide adolescents toward informed, interest-aligned decisions can have a lasting impact on their futures. This research is grounded in Holland's (1997) influential model of vocational interests, which posits that individuals are more likely to thrive and find fulfillment in environments that align with their personal interests. Holland's model identifies six distinct dimensions—Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC)—which classify both individual interests and the characteristics of their environments. This framework enables a consistent and meaningful measurement of the fit between individuals and their educational or vocational settings.

Chapter 5 adopts a longitudinal perspective, exploring the development of PE interest fit during the first three years of secondary education. Given that students' vocational interests are likely still developing at this stage, a longitudinal study design was employed to examine how interests and the interest fit evolve over time. According to Holland’s (1997) theory of vocational interests, individuals seek environments that match their evolving interests, suggesting that PE fit improves as individuals gravitate toward environments that reflect their changing preferences. Similarly, Schneider’s (1987) Attraction-Selection-Attrition (ASA) model further underscores the dynamic nature of PE fit, highlighting how individuals are drawn to environments that match their personal attributes, and over time, select or exit situations based on how well these environments align with their evolving characteristics. While existing research has demonstrated the importance of both selection (Hanna et al., 2021; Roberts & Robins, 2004; Wille & De Fruyt, 2014) and socialization (Ertl et al., 2023; Nye et al., 2021) in shaping PE fit during late adolescence and adulthood, there is limited understanding of how these processes unfold during early adolescence—a period marked by shifts in interests and pivotal educational choices (Slot et al., 2021). This study seeks to fill that gap by exploring the longitudinal patterns of PE interest fit during early adolescence and aims to improve academic counseling practices in secondary education. Given the well-documented gender differences in vocational interests throughout adolescence and adulthood, along with inconsistent findings regarding developmental patterns, gender is an important control variable in this study. By accounting for these differences, we offer a more nuanced understanding of how vocational interests evolve, which strengthens the validity and robustness of our findings.

Finally, chapter 6 brings together the findings from the previous chapters and explores their practical implications for enhancing student guidance. This chapter also outlines directions for future research and proposes strategies for the continued refinement and expansion of the Talentcenter model.

References

Anttila, S., Lindfors, H., Hirvonen, R., Määttä, S., & Kiuru, N. (2023). Dropout intentions in secondary education: Student temperament and achievement motivation as antecedents. Journal of Adolescence, 95, 248-263. https://doi.org/10.1002/jad.12110

Aschenbrenner, S., Kaiser, S., Pfüller U., Roesch-Ely, D., & Weisbrod, M. (2012). *Manual Cognitive Basic Testing* (Version 2). Schuhfried

Ball, S. J. (2003). Class strategies and the educational marketplace: The middle-classes and social advantage. London: Routledge

Bartlett, K. A., & Camba, J. D. (2023). The role of a graphical interpretation factor in the assessment of spatial visualization: A critical analysis. *Spatial Cognition and Computation, 23,* 1–30. https://doi.org/10.1080/13875868.2021.2019260

Bartlett, K. A., Palacios-Ibáñez, A., & Camba, J. D. (2024). Design and Validation of a Virtual Reality Mental Rotation Test. *ACM Transactions on Applied Perception, 21,* 5. https://doi.org/10.1145/3626238

Bourdeaud'hui, H., Aesaert, K., Van Braak, J. (2020). Identifying student- and class-level correlates of sixth-grade students' listening comprehension. *L1-educational studies in language and literature, 20,* 1-38. https://doi.org/10.17239/L1ESLL-2020.20.01.04

Brysbaert, M., Keuleers, E., Mandera; P., & Stevens, M. (2014). *Woordenkennis van Nederlanders en Vlamingen anno 2013: Resultaten van het Groot Nationaal Onderzoek Taal* [Vocabulary knowledge of Dutch and Flemish speakers in 2013: Results of the 'Groot Nationaal Onderzoek Taal.']. Ghent, Belgium: Academia Press.

Cheng, Y. L., & Mix, K. S. (2014). Spatial Training Improves Children's Mathematics Ability. *Journal of Cognition and Development*, *15*, 2–11. https://doi.org/ 10.1080/15248372.2012.725186

Condon, D., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Credé, M., & Kuncel, N. (2008). Study habits, skills, and attitudes. The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3*, 425-453. https://doi.org/10.1111/j.1745-6924.2008.00089.x

Dawis, R. V. (2005). The Minnesota Theory of Work Adjustment. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 3-23). Hoboken, NJ: John Wiley & Sons Inc.

Degroote, E., Demanet, J., & Van Houtte, M. (2020) The influence of school mobility and dropout rates on non-mobile students' school engagement: a chicken-and-egg situation? *Research Papers in Education, 35,* 443-466, https://doi.org/10.1080/02671522.2019.1601755

Department of Education and Training. (2023). *Voltijds gewoon secundair onderwijs: algemeen overzicht* [Full-time regular secondary education: General overview] [Dataset]. Statistical yearbook of Flemish education 2023-2024. Flemish Government. https://onderwijs.vlaanderen.be/nl/onderwijsstatistieken/statistisch-jaarboek/statistisch-jaarboek-van-het-vlaams-onderwijs-2023-2024

Desoete, A., & Roeyers, H. (2006). Cognitieve Deelhandelingen van het Rekenen (CDR). *Handleiding & testprotocol [Cognitive Developmental skills in Arithmetics. Manual & testprotocol].* Herenthals: VVL.

de Vries, N., Meeter, M., & Huizinga, M. (2024). Does interest fit between student and study program lead to better outcomes? A meta-analysis of vocational interest congruence as predictor for academic success. *Educational Research Review, 44,* 100619. https://doi.org/10.1016/j.edurev.2024.100619

De Windt, H. (2020, January, 9). 18.000 middelbare scholieren veranderden tijdens schooljaar van richting: "Watervalsysteem bestaat nog altijd". *Vrt nws.* Retrieved January 21, 2025, from https://www.vrt.be/vrtnws/nl/2020/01/09/18-000-middelbare-scholieren-veranderden-tijdens-schooljaar-van/

De Witte, K., Cabus, S., Thyssen, G., Groot, W., & Maassen van den Brink, H. (2013). A critical review of the literature on school dropout. *Educational Research Review, 10,* 13-28, https://doi.org/10.1016/j.edurev.2013.05.002

Draelants, H. (2014). Des héritiers aux initiés? Note sur les nouvelles modalités de la reproduction sociale par l'école. *Social Science Information, 53,* 403–432. https://doi.org/ 10.1177/0539018414522893

Dutry, M., Vereeck, A., Duyck, W., Derous, E., Schelfhout, S., Szmalec, A., … Dirix, N. (2025). Validation of the Children's International Cognitive Ability Resource (Ch-

ICAR). *Behaviour Research Methods, 57.* https://doi.org/10.3758/s13428-024-02591-1.

Dutry, M., Schelfhout, S., Derous, E., Duyck, W., & Dirix, N. (n.d.). Optimizing career guidance in secondary education: Exploring person-environment interest fit with the SIMON-jr interest inventory. *Manuscript in Preparation.*

Ertl, B., Hartmann, F. G., & Wunderlich, A. (2023). Stability of vocational interests during university studies. *Journal of Individual Differences, 44, 143*-162. https://doi.org/10.1027/1614-0001/a000392

Fall, M. A., & Roberts, G. (2012). High School Dropouts: Interactions between Social Context, Self-Perceptions, School Engagement, and Student Dropout. *Journal of Adolescence, 35,* 787–798. https://doi.org/10.1016/j.adolescence.2011.11.004.

Flanagan, D. P., & McDonough, E. M. (2018). *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (Fourth Edition). Guilford Press.

Flemish government. (n.d.-1). *Onderwijs in Vlaanderen* [Education in Flanders]. Retrieved January 14, 2025, from Onderwijs.vlaanderen.be. https://onderwijs.vlaanderen.be/

Flemish government. (n.d.-2). *Evaluaties in onderwijsinstellingen: Secundair onderwijs*. Retrieved January 14, 2025, from https://www.vlaanderen.be/onderwijs-en-vorming/evaluaties-in-onderwijsinstellingen/secundair-onderwijs

Flemish government. (n.d.-3). *Vlaamse toetsen*. Retrieved January 14, 2025, from https://www.vlaanderen.be/onderwijs-en-vorming/vlaamse-toetsen

Flum, H., & Blustein, D. L. (2000). Reinvigorating the study of vocational exploration: A framework for research. *Journal of Vocational Behavior, 56,* 380-404, https://doi.org/10.1006/jvbe.2000.1721

Fonteyne, L. (2017). *Constructing SIMON : a tool for evaluating personal interests and capacities to choose a post-secondary major that maximally suits the potential.* Ghent University. Faculty of Psychology and Educational Sciences.

Fonteyne, L., Duyck, W., & De Fruyt, F. (2017). Program-specific prediction of academic achievement on the basis of cognitive and non-cognitive factors. *Learning and Individual Differences, 56,* 34-48. https://doi.org/ 10.1016/j.lindif.2017.05.003

Goosen, K., Boone, S., Dehertogh, B., Kavadias, D., Mahieu, P., Van Avermaet, P., … Van Kerckhove, C. (2017). *Is dat iets voor mij juf? Leerlingen versterken in het keuzeproces van basis naar secundair.* Leuven: LannooCampus.

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology, 28,* 545–579. https://doi.org/10.1037/0022-0167.28.6.545

Gottfredson, L. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132. https://doi.org/10.1016/S0160-2896(97)90014-3

Griep, Y., Kinnunen, U., Nätti, J., De Cuyper, N., Mauno, S., Mäkikangas, A., & De Witte, H. (2016). The effects of unemployment and perceived job insecurity: A comparison of their association with psychological and somatic complaints, self-rated health and life satisfaction. *International Archives of Occupational and Environmental Health*, 89, 147–162. https://doi.org/10.1007/s00420-015-1059-5

Hanna, A., Briley, D., Einarsdóttir, S., Hoff, K., Rounds, J. (2021). Fit gets better: A longitudinal study of changes in interest fit in educational and work environments. *European Journal of Personality, 35,* 557-580. https://doi.org/ 10.1177/08902070211014022

Hanna, A., & Rounds, J. (2020). How accurate are interest inventories? A quantitative review of career choice hit rates. *Psychological Bulletin, 146,* 765–796. https://doi.org/10.1037/bul0000269

Hegarty, M., & Waller, D. A. (2005). Individual Differences in Spatial Abilities. In P. Shah & A. Miyake (Eds.), The Cambridge Handbook of Visuospatial Thinking (121–170). Cambridge University Press.

Hirschi, A. (2009). Career adaptability development in adolescence: Multiple predictors and effect on sense of power and life satisfaction. *Journal of Vocational Behavior, 74,* 145–155. https://doi.org/10.1016/j.jvb.2009.01.002.

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL, US: Psychological Assessment Resources.

Janosz, M., I. Archambault, J. Morizot, & Pagani, L. S. (2008). School engagement trajectories and their differential predictive relations to dropout. *Journal of Social Issues, 64,* 21–40. https://doi.org/10.1111/josi.2008.64.issue-1

Jiang, Z., Newman, A., Le, H., Presbitero, A., & Zheng, C. N. (2019). Career exploration: A review and future research agenda. *Journal of Vocational Behavior, 110,* 338-356. https://doi.org/ 10.1016/j.jvb.2018.08.008

Kalejaiye, P.O. (2023). Addressing shortage of skilled technical workers in the USA: A glimpse for training service providers. *Future Business Journal, 9,* 50. https://doi.org/10.1186/s43093-023-00228-x

Lee, I. J. (2020). Using augmented reality to train students to visualize three-dimensional drawings of mortise-tenon joints in furniture carpentry. *Interactive Learning Environments, 28,* 930-944. https://doi.org/ 10.1080/10494820.2019.1572629

Lee, V.E., & Burkam, D.T. (2003). Dropping out of high school: the role of school organization and structure. *American Educational Research Journal, 40,* 353-393, https://doi.org/10.3102/00028312040002353

Lent, R. W., & Brown, S. D. (2013). Social cognitive model of career self-management: Toward a unifying view of adaptive career behavior across the life span. *Journal of Counseling Psychology, 60,* 557–568. https://doi.org/10.1037/a0033446.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498. https://doi.org/10.2307/1130467

Lipnevich, A. A., & Roberts, R. D. (2012). Noncognitive skills in education: Emerging research and applications in a variety of international contexts. *Learning and Individual Differences, 22*, https://doi.org/173-177. doi:10.1016/j.lindif.2011.11.016

Lochhead, I., Hedley, N., Çöltekin, A., & Fisher, B. (2022). The Immersive Mental Rotations Test: Evaluating Spatial Ability in Virtual Reality. *Frontiers in Virtual Reality, 3,* 5. https://doi.org/10.3389/frvir.2022.820237

Neureiter, M., & Traut-Mattausch, E. (2017). Two sides of the career resources coin: Career adaptability resources and the impostor phenomenon. *Journal of Vocational Behavior, 98,* 56–69. https://doi.org/10.1016/j.jvb.2016.10.002

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7,* 384–403. https://doi.org/10.1177/1745691612449021

Nye, C. D., Wille, B., Amory, J., & De Fruyt, F. (2021). Are work activities related to interest change over time? A 22-year longitudinal study. J*ournal of Personality and Social Psychology, 121*, 865–893. https://doi.org/10.1037/pspp0000360

OECD. (2020). *Education at a Glance 2020: OECD Indicators.* OECD Publishing. https://doi.org/10.1787/69096873-en

Piri, Z., & Cagiltay, K. (2024). Can 3-Dimensional Visualization Enhance Mental Rotation (MR) Ability?: A Systematic Review. *International Journal of Human-Computer Interaction, 40,* 3683-3698. https://doi.org/ 10.1080/10447318.2023.2196161

Ramsdal, G. H., Bergvik, S., & Wynn, R. (2018). Long-term dropout from school and work and mental health in young adults in Norway: A qualitative interview-based study. *Cogent Psychology, 5.* https://doi.org/10.1080/23311908.2018.1455365

Revelle, W., Condon, D., Wilt, J., French, J., Brown, A., & Elleman, L. (2017). Web- and phone-based data collection using planned missing designs. In N. Fielding, R. Lee, & G. Blank (Eds.), *The SAGE Handbook of Online Research Methods* (2nd ed., pp. 578–595). Sage Publishing.

Revelle, W., Dworak, E., & Condon, D. (2020). Cognitive ability in everyday life: The utility of open-source measures. *Current Directions in Psychological Science*, *29*(4), 358–363. https://doi.org/10.1177/0963721420922178

Roberts, B. W., & Robins, R. W. (2004). Person-environment fit and its implications for personality development: A longitudinal study. *Journal of Personality*, *72*, 89-110. https://doi.org/ 10.1111/j.0022-3506.2004.00257.x

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53,* 118-137. https://doi.org/10.1016/j.intell.2015.09.002

Rumberger, R. W., & Rotermund, S. (2012). The relationship between engagement and high school dropout. In *Handbook of Research on Student Engagement*, edited by S. L. Christenson, A. L. Reschly, and C. Wylie, 491–513. New York, NY: Springer.

Ryan, R. M., & Connell, J. P. (1989). Perceived locus of causality and internalization: Examining reasons for acting in two domains. *Journal of Personality and Social Psychology, 57,* 749-761. https://doi.org/10.1037/0022-3514.57.5.749

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102,* 254–273. https://doi.org/10.1037/apl0000151

Savickas, M. L. (2002). Career construction: A developmental theory of vocational behavior. In D. Brown (Ed.). *Career choice and development* (pp. 255–311). (4th ed.). San Francisco, CA: Jossey-Bass.

Savickas, M. L. (2013). Career construction theory and practice. In S. D. Brown, & R. W. Lent (Eds.). *Career development and counseling: Putting theory and research to work* (pp. 147–183). (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Slot, E. M., Bronkhorst, L. H., Akkerman, S. F., & Wubbels, T. (2021). Vocational interest profiles in secondary school: Accounting for multiplicity and exploring associations with future-oriented choices. *Journal of Educational Psychology, 113,* 1059-1071. https://doi.org/10.1037/edu0000475

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., Derous, E., De Fruyt, F., & Duyck, W. (2021). How interest fit relates to STEM study choice: Female students fit their choices better. *Journal of Vocational Behavior, 129,* 103614. https://doi.org/10.1016/j.jvb.2021.103614

Schmidt, F. L. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings.* ResearchGate. Available at https://doi.org/10.13140/RG.2.2.18843.26400

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437-453. doi:10.1111/j.1744-6570.1987.tb00609.x

Schneider, W., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). The Guilford Press.

Schoppe, K. J. (1974). Das MLS-Gerät: Ein neuer Testapparat zur Messung feinmotorischer Leistungen. *Diagnostica, 20.* 43–46

Schuhfried (2018). *Vienna Test System* (Version 8.11) [Computer software]. Author.

Seghers, M., Boone, S., & Van Avermaet, P. (2019). Classed patterns in the course and outcome of parent-teacher interactions regarding educational decision-making. *Educational Review, 73,* 417-435. https://doi.org/10.1080/00131911.2019.1662771

Sins, P., de Leeuw, R., de Brouwer, J., & Vrieling-Teunter, E. (2024). Promoting explicit instruction of strategies for self-regulated learning: evaluating a teacher professional development program in primary education. *Metacognition Learning 19,* 215–247. https://doi.org/10.1007/s11409-023-09368-5

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences, 26,* 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Spruyt, B. (2014). De onderwaardering van technisch en beroepsonderwijs: oorsprong en gevolgen. In I. Nicaise, B. Spruyt, M. Van Houtte & D. Kavadias (Eds.). *Het onderwijsdebat. Waarom de hervorming van het secundair onderwijs broodnodig is* (pp. 71-88). Antwerpen: EPO.

Statistcs Flanders. (2024). *Vroegtijdige schoolverlaters op basis van administratieve data.* Retrieved January 15, 2025, from https://www.vlaanderen.be/statistiek-vlaanderen/onderwijs-en-vorming/vroegtijdige-schoolverlaters-op-basis-van-administratieve-data

Stevens, P. A. J. (2007). Exploring the importance of teachers' institutional structure on the development of teachers' standard. *Sociology of Education, 80,* 314-329. https://doi.org/10.1177/003804070708000402

Su, R. (2020). The three faces of interests: An integrative review of interest research in vocational, organizational, and educational psychology. *Journal of Vocational Behavior, 116,* 103240. https://doi.org/10.1016/j.jvb.2018.10.016.

Tikhomirova, T., Malykh, A., & Malykh, S. (2020). Predicting academic achievement with cognitive abilities: Cross-sectional study across school education. *Behavioral Sciences, 10:*158. https://doi.org/10.3390/bs10100158

Uttal, D. H., McKee, K., Simms, N., Hegarty, M., & Newcombe, N. S. (2024). How Can We Best Assess Spatial Skills? Practical and Conceptual Challenges. *Journal of Intelligence, 12,* 8. https://doi.org/10.3390/ jintelligence12010008

Vandenbroeck, M., Dockx, J., & Janssen, R. (2024). Grade 12 students' perceptions of educational tracks in Flanders. *Social Psychology Of Education, 27,* 1065–1095. https://doi.org/10.1007/s11218-024-09901-7

Vandevelde, S., Van Keer, H. & Rosseel, Y. (2013). Measuring the complexity of upper primary school children's self-regulated learning: A multi-component approach. *Contemporary educational Psychology, 38,* 407-425. https://doi.org/10.1016/j.cedpsych.2013.09.002

Van Houtte, M., & Demanet, J. (2016). Teachers' beliefs about students, and the intention of students to drop out of secondary education in Flanders. *Teaching and Teacher Education, 54,* 117-127. https://doi.org/10.1016/j.tate.2015.12.003

Van Praag, L., Boone, S., Stevens, P., & Van Houtte, M. (2015). De paradox van het watervalsysteem: wanneer het groeperen van studenten in homogene groepen tot meer heterogeniteit leidt in het beroepsonderwijs. *Sociologis, 36,* 82-101.

Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., & Lens, W. (2009). Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of Educational Psychology, 101*, 671-688. https://doi.org/10.1037/a0015083

Vantieghem, W., & Van Houtte, M. (2015). Are Girls more Resilient to Gender-Conformity Pressure? The Association Between Gender-Conformity Pressure and Academic Self-Efficacy. *Sex roles, 73,* 1–15. https://doi.org/10.1007/s11199-015-0509-6

Verhaeghe, O., & Lamote, S. (2025, January 16). Technisch en beroepssecundair onderwijs populairder dan ooit. *De Tijd.* Retrieved January 17, 2025, from https://www.buurtsuper.be/nl/nieuws-uit-de-sector/meerderheid-volgt-momenteel-tso-of-bso-opleiding

Veurink, N. L., & Sorby, S. A. (2019). Longitudinal study of the impact of requiring training for students with initially weak spatial skills. *European Journal of Engineering Education, 44,* 153–163. https://doi.org/10.1080/03043797.2017.1390547

Voka. (2024). *Aantal schoolverlaters zonder diploma nog nooit zo hoog.* Retrieved January 15, 2025, from https://www.voka.be/nieuws/aantal-schoolverlaters-zonder-diploma-nog-nooit-zo-hoog

Wille, B., & De Fruyt, F. (2014). Vocations as a source of identity: Reciprocal relations between Big Five personality traits and RIASEC characteristics over 15 years. *Journal of Applied Psychology*, *99*, 262. https://doi.org/10.1037/a0034917

Zheleva, A., De Bruyne, J., Durnez, W., Van De Walle, S., Verreyken, S., Demanet, J., & Bombeke, K. (2023). CaliBrainVR: Using psycho-physiological measures to calibrate virtual reality training. *2023 15th International Conference on Quality of Multimedia e-Experience, QOMEX,* 123-126. https://doi.org/10.1109/QOMEX58391.2023.10178522

Chapter 2.

Validation of the Children's International Cognitive Ability Resource (Ch-ICAR)

Dutry, M., Vereeck, A., Duyck, W., Derous, E., Schelfhout, S., Szmalec, A., Woumans, E., Schittekatte, M., Debeer, D., & Dirix, N. (2025). Validation of the Children's International Cognitive Ability Resource (Ch-ICAR). Behavior Research Methods, 57. https://doi.org/ 10.3758/s13428-024-02591-1

Abstract

The International Cognitive Ability Resource, abbreviated ICAR, counters some of the practical problems researchers face when using good, but proprietary licensed intelligence tests like the Wechsler tests, which include unfeasible administration times and financial costs. So far, ICAR has been validated for adolescents and adults in many countries, offering a viable test alternative for these populations. For use among children, however, the appropriateness of this resource was yet unknown. Therefore, we set out to develop a children's ICAR: an instrument composed of ICAR-items, which provides a measure of cognitive ability in children between 11 and 14 years of age. The present article discusses the compilation process of the Ch-ICAR drawing from a pilot study, and evaluates its validity based on two additional studies. The pilot study involved 99 primary school pupils and aimed to select items for the Ch-ICAR instrument. Study 1 investigated the basic psychometric qualities of the Ch-ICAR in a sample of 820 secondary school pupils. Study 2 examined the construct validity by cross-validating the Ch-ICAR with on the one hand Raven's 2 Progressive Matrices, and on the other hand the Flemish CoVaT-CHC Basic Version, relying on samples of 91 secondary and 96 primary school pupils respectively. Results support the utility of the Ch-ICAR as a measure of children's cognitive abilities within a research context.

Introduction

"[L]ife is an intelligence test, and the importance of measuring intellectual ability as people respond to life's challenges cannot be overstated." (Dworak et al., 2021)

Cognitive ability refers to the capacity to acquire, process, manipulate, organize and apply knowledge effectively (Gottfredson, 1997). It is widely recognized to reliably predict important outcomes such as academic achievement (Roth et al., 2015; Tikhomirova et al., 2020; Vilia et al., 2017) and work performance (Sackett et al., 2017; Schmidt, 2016). Furthermore, cognitive ability is significantly associated with happiness (Ali et al., 2013; Kanazawa, 2013) and overall health (Wrulich et al., 2014). Unsurprisingly, cognitive ability and its assessment therefore belong to the most discussed and central topics within psychology (Gottfredson & Saklofske, 2009).

For children, the most commonly used test worldwide to measure cognitive ability is the Wechsler Intelligence Scale for Children (WISC; Evers et al., 2012; Flanagan & Alfonso, 2017; Weiss et al., 2019). The WISC is a time- and field-tested instrument that has strong psychometric features (Oakland et al., 2016) and boasts international positive reviews (e.g., Na & Burns, 2016 for the English version; Kwaliteitscentrum voor Diagnostiek, 2023 for the Dutch version). However, the WISC is primarily designed for clinical contexts, and less attuned to the needs of many research settings. Because this test can only be administered individually, the workload soon becomes unmanageable for studies aiming at larger samples, even when using abbreviated scales like the WASI-II (Wechsler, 2011) or other short forms containing a reduced number of subtests (Aubry & Bourdin, 2018). Financially as well, the cost of administering the WISC often surpasses budgets of (junior) researchers.

To alleviate the practical problems that researchers encounter with Wechsler Scales and other proprietary licensed intelligence tests, and "to address the need for psychometrically valid tools that are well-suited for large-scale, remote data collection" (Dworak et al., 2021, p.

3), an interuniversity team of cognitive ability experts developed the International Cognitive Ability Resource (ICAR; Revelle et al., 2020), a public domain measure of cognitive ability designed for online administration.

Public domain measures are valuable tools in a research context. Open access instruments enable more researchers to conduct studies, share findings, and ultimately advance the field (Condon & Revelle, 2014). The public nature of these instruments encourages collaboration within the research community regarding test development, refinement, and validation (Goldberg, 1999). On the downside, publicly available measures of cognitive ability sometimes either lack sufficient psychometric rigor (e.g., Cog15; Kajonius, 2014; Kristjánsdóttir & Zaiter, 2023) or are too narrow in focus, concentrating exclusively on a particular type of item (e.g., only matrix reasoning; Zorowitz et al., 2023).

The ICAR, on the other hand, is psychometrically sound and offers a wide variety of subtests (Revelle et al., 2017) including items of high and low difficulty. The items can be programmed into any software of choice, and the extensive database (The International Cognitive Ability Resource Team, 2014) makes it possible to tailor the instrument to the intended audience or to switch up the items and minimize test-retest effects. In the earliest days of the project, the ICAR consisted of just 60 items, divided over four subtests: Verbal Reasoning, Three-dimensional Rotation, Matrix Reasoning, and Letter-Number Series (Condon & Revelle, 2014). This original ICAR60, as well as a 16-item subset known as the ICAR16, were validated in a sample of 96,958 participants with an age range from 14 to 90 years (Condon & Revelle, 2014). Since its initial validation, the ICAR has established itself as a cognitive ability measure for research on adolescents and adults (used in 79 published studies according to Dworak et al., 2021, and more since), the majority of studies employing the ICAR16.

The appropriateness of an ICAR for children, however, is still unknown. Considering the dynamic nature of brain development across the lifespan, with rapid changes in early childhood and more gradual changes in adulthood, it is essential to acknowledge that the ICAR items designed for adults may not be adequate for children (Flanagan & McDonough, 2018). For instance, through neurodevelopment, brains become more organized, efficient, and faster at processing information during childhood and adolescence (Lenroot & Giedd, 2006; Low & Cheng, 2006). Moreover, research suggests that age and frontal lobe development play pivotal roles in shaping cognitive ability (Shaw et al., 2006), highlighting the need for assessment tools tailored to the cognitive abilities of children. Therefore, we aimed to compile an instrument based on ICAR items that unites the existing practical advantages of the ICAR with an adequate measurement of the cognitive ability of children between the ages of 11 and 14. As the ICAR project is not based a priori on a particular theoretical framework such as the Cattell-Horn-Carroll (or CHC-) model of intelligence (Schneider & McGrew, 2012), we also looked into the question which broad cognitive abilities are measured by which ICAR subtest. The compilation of a children's ICAR, or Ch-ICAR in brief, not only broadens the scope of the ICAR project: It also addresses a notable gap in the literature, for at the moment there is a lack of psychometrically sound measures of cognitive ability that are freely available, limited in administration time and suited for children.

In this article, we first discuss the pilot study we carried out to compile the Ch-ICAR. Next, we discuss two studies conducted to assess the psychometrics of this newly compiled instrument. Study 1 investigated the basic psychometric qualities of the Ch-ICAR. Study 2 cross-validated the Ch-ICAR with two commercial tests, on the one hand Raven's 2 Progressive Matrices, and on the other hand the Flemish CoVaT-CHC Basic Version. The former is a measure of nonverbal intelligence, which provides an effective estimate of cognitive ability (McLeod & McCrimmon, 2021) and is commonly used and suitable for group administration (Oakland et al., 2016). The latter is a comprehensive measure of intelligence, whose composition reflects the CHC-model of intelligence (Magez et al., 2015). In Flanders, the CoVaT offers a valid alternative to other intelligence tests, since − unlike Raven's Matrices − it encompasses multiple broad cognitive abilities (Flanagan & McDonough, 2018). Cross-validation of the Ch-ICAR using both Raven and CoVaT instruments provides us with empirical verification about whether the Ch-ICAR is indeed a valid measure for children's cognitive ability.

Pilot study

In an initial phase, we solicited feedback from experts in the field and from children with the target age to assess the suitability of various ICAR item types[4] for constructing a preliminary instrument. This preliminary instrument consisted of four subtests: Verbal Reasoning, Matrix Reasoning, Number Series, and Figural Analogies, with 16, 11, 23 and 20 items respectively (we adopted all items available on the ICAR-website, except for the subtest Number Series, where we selected a mix of easy and more challenging items). Verbal Reasoning contains questions on vocabulary and general knowledge as well as mathematical problems[5]. Matrix Reasoning items consist of eight geometric shapes, placed in a 3-by-3 grid, with one shape missing; participants have to choose the response option that represents the missing ninth shape. Number Series items show a short sequence of numbers; participants have to complete the sequence by entering the next number of the sequence. Figural Analogies items consist of three colored geometric shapes, with the first two shapes standing in a certain relation to each other; participants have to choose the response option that

[4]Including Verbal Reasoning, Three-dimensional Rotation, Matrix Reasoning, Letter-Number Series, Number Series and Figural Analogies. We decided not to include Three-dimensional Rotation and Letter-Number Series in the preliminary instrument because the former proved unsuitable for the target audience, with half of the children unable to solve any items, and the latter raised concerns about test conditions—specifically, if participants are allowed (or not prevented) from writing down the alphabet, the items become significantly easier.

[5]As such, in our opinion the name of the subtest does not quite cover the content, but for matters of consistency and comparability, we chose to retain the original subtest name.

represents the missing fourth shape, which stands in the same relation to the third as the second does to the first. A specimen of each item type can be found in Table 2A1 (the final compilation of the Ch-ICAR is discussed further and is available on the ICAR website[6]).

Subsequently, we carried out a pilot study with the preliminary instrument. The main objective of this pilot study was to select items for the Ch-ICAR instrument that provide an accurate estimate of cognitive ability, while putting minimal strain on researchers' time and participants' cognitive load. With that goal in mind, we aimed to reduce the number of items to ten at most per subtest.

Five class groups of three different primary schools in Flanders[7] took part in the pilot study (N_{Pilot} = 99, male 41%). All participants were in their last year of primary education (age 11-12). In Flanders, primary education is marked by its comprehensive nature, the absence of ability groups, and open and free accessibility. As a result, Flemish primary education classrooms exhibit a diverse and inclusive pupil population. For practical reasons, each class only took two or three of the four subtests we had previously selected. The combination and order of subtests were assigned at random, the items were administered digitally, and testing took place in the classroom. In line with the recommendation of the ICAR team to take into account possible interruptions due to device malfunctions (Condon & Revelle, 2014), no time limit was imposed.

We noticed that many participants asked for additional explanation about the subtests Matrix Reasoning and Figural Analogies during the test-taking process. To improve the instructions in the final instrument, we added an example item to each of these two subtests. Seeing that items of limited complexity were scarce in the ICAR database, we designed the

[6]The Ch-ICAR can be accessed via this link: https://icar-project.com/projects/childrens-icar, users will only need to create a user account.

[7]For the Verbal Reasoning items, we procured the Dutch translation by Fontaine (Nelissen et al., In preparation). With regard to the instructions, we provided each subtest with a written Dutch explanation ourselves.

example items ourselves (Table 2A2). Upon completing an example item, participants receive immediate feedback regarding the correctness of the chosen answer and an explanation of why the chosen answer is either correct or incorrect.

For the final item selection, we prioritized items with stronger item-subscale correlations while maintaining a balanced range of difficulty levels. Additionally, we ensured that selected items showed minimal sex differences and avoided disproportionate response times. This approach allowed us to build a diverse and robust final set of items for the Ch-ICAR. Detailed statistics for each subtest can be found in Appendix 2B. Ultimately, we selected a total of 31 items for the final Ch-ICAR: 8 items for Verbal Reasoning, Number Series and Matrix Reasoning each, and 7 for Figural Analogies (Appendix 2C contains the final item set). Table 1 provides a concise overview of the primary findings of the pilot study. This table serves as a brief summary of key insights and observations from the preliminary study. For a more comprehensive understanding and detailed item-level information, readers are invited to consult Appendix 2B.

Table 1

Primary Findings of the Pilot Study

Subtest	*N*	Number of items in preliminary instrument	Cronbach's alpha in preliminary instrument	Average PC (*SD*) in preliminary instrument	Cronbach's alpha in final Ch-ICAR	Average PC (*SD*) in final Ch-ICAR
Verbal Reasoning	46	16	.50	.34 (.14)	.65	.40 (.24)
Matrix Reasoning	66	11	.63	.29 (.20)	.69	.31 (.25)
Number Series	71	23	.88	.35 (.20)	.85	.43 (.31)

Figural Analogies	56	20	.77	.23 (.18)	.83	.33 (.31)

Note. Average PC is the average proportion of correct responses based on the total score of the participants that completed the full subtest.

Study 1: Basic psychometric qualities of the Children's ICAR

Study 1 investigated the basic psychometric qualities of the Ch-ICAR in a large sample of pupils in Flanders. We assessed (1) the internal structure of the test and the psychometric properties of the individual items, (2) the distribution of the Ch-ICAR scores and the distribution of the Ch-ICAR scores over sex (although small sex differences in specific cognitive abilities may exist, the equality in cognitive ability is well established in literature (e.g., Deary et al., 2007; Fergusson & Horwood, 1997; Nisbett et al., 2012)), (3) the internal consistency of the different subtests and the Ch-ICAR as a whole, and (4) the construct validity by examining the relation with school achievement and a basic mathematical skill test.

Method

Participants

A sample of Flemish pupils (N_1 = 820, male 53.5%, female 45.7%, .8% did not specify their sex) was recruited in 23 different schools all across Flanders. All participants were in their first year of regular secondary education (M_{age} = 12.02, SD_{age} = .56) and other than that there were no exclusion criteria for participation. Table 2 offers a comprehensive overview of the demographics and associated statistics.

The research project[8] encompassing the ICAR studies was approved by the Ethical Commission of the Faculty of Psychology and Educational Sciences of Ghent University (reference number 2021/59). Both the parents of the participants and the participants

[8]Entitled: "Study orientation in secondary education".

themselves gave their informed consent for participation, as well as for storage and use of the data by the researchers.

Table 2

Demographic Information Study 1 and Study 2

	Study 1	Study 2 (RPM sample)	Study 2 (CoVaT sample)
N	820	91	96
N female (%)	375 (46%)	55 (60%)	48 (50%)
Mean age (*SD*) (years)	12.02 (.56)	12.99 (.57)	11.05 (.27)
Age range (years)	10 - 14	12 - 15	10 - 12
Socio-economic status			
N home language is not Dutch (%)	83 (10%)	3 (3%)	3 (3%)
*N r*eceiving a bursary (%)	231 (28%)		
N highest attained degree is primary education or lower Parent 1 (%)	83 (10%)		
N highest attained degree is primary education or lower Parent 2 (%)	81 (10%)		
N schools	23	2	3
M number of pupils per school *(SD)*	35.65 (22.95)	45.50 (38.89)	32.00 (23.64)
Number of pupils per school: range	8- 83	18 - 73	15 - 59

Materials

Socio-economic status. To define the socio-economic status (SES) of the participants, we used a questionnaire for parents with three indicators that are official markers for school funding in Flanders: financial capacity of the parents, educational level of both parents, and home language (Ministry of Education and Training, 2012). Following official criteria for these indicators, the following participant characteristics were seen as indicative for low SES: receiving a bursary, having one or two parents whose highest attained degree is primary education or lower, and having another home language than Dutch. In our sample, 10% of the participants had another home language than Dutch (for 8% home language was unspecified), 28% of the parents received a bursary (for 17% financial capacity was unspecified), 10% had one parent with primary education or lower as highest degree and 5% had both parents with primary education or lower as highest degree (for 9% educational level of the parents was unspecified).

Children's ICAR. To recapitulate, the Ch-ICAR as compiled after our pilot study comprises eight Verbal Reasoning items, eight Matrix Reasoning items, eight Number Series items, and seven Figural Analogies items (see Pilot study for a description of the subtests). Except for Number Series items (which is open-ended), each item consists of eight response options, including *'none of these'* and *'I don't know'*. Responses on the items are converted to binary values: Right answers are scored as one, wrong answers as zero. The subtest scale scores are calculated as the sum of the binary values of each item of the subtest. The Ch-ICAR total score is calculated as the sum of the subtest scale scores.

Academic performance. ***School achievement.*** Data on school achievement include the GPA (weighted average of assignments and tests of all subjects on a scale from 0 to 100), as well as four separate subjects scores: Mathematics, Dutch, Natural Sciences, and Technics (also on a scale from 0 to 100). As secondary education in Flanders does not feature standardized tests at the time of writing, the GPA and subject scores are based on the assessment of the individual teachers. To avoid possible bias, we standardized the achievement scores within schools[9]. Schools with less than ten participants were excluded for the corresponding analyses as well as pupils that changed schools within the year. This resulted in a sample of 722 pupils, drawn from a total of 19 distinct schools.

Basic mathematical skill test: CDR. We included an additional, standardized external criterion, related to cognitive ability, to detect and control for school specific effects. We administered a basic mathematical skill test to all participants, namely the shortened version (18 items, McDonald's omega (ω) = .72) of the *Cognitive Developmental Skills in Arithmetics* (*Cognitieve Deelhandelingen van het Rekenen* [CDR]; Desoete & Roeyers, 2006).

Procedure

Participants completed the Ch-ICAR and the CDR at the beginning of the school year (September-November 2021). The instruments were administered digitally, in the classroom of the participants, during school hours. The pupils' teacher as well as a university supervisor were present. The tests were programmed in Qualtrics (https://www.qualtrics.com), a commercial survey software and research platform.

After a general introduction, participants could work in silence at their own pace on the different tests. The subtests of the Ch-ICAR appeared in a fixed sequence: Verbal Reasoning, Matrix Reasoning, Number Series, and Figural Analogies. The items appeared one by one, in ascending difficulty (as defined in the pilot study). Once participants decided

[9]The first year of secondary education in Flanders consists of two primary streams: 1A and 1B (Ministry of Education and Training, s.d.). Admission to 1A is automatic for those who possess a certificate of primary education, while entry into 1B is automatic for those without the certificate. However, pupils holding a certificate of primary education may gain admission to 1B through a favorable decision of the admissions class council and the student guidance center, and with parental agreement. Consequently, the predominant majority of pupils embark on the 1A track (87% in the academic year 2021-2022; Ministry of Education and Training, 2022). As pupils in 1B follow a distinct curriculum, we opted for the separate calculation of standardized scores for this subgroup within the school. B-groups with fewer than 10 pupils were also excluded from the corresponding analyses, resulting in the exclusion of 6 B-groups.

to move on to the next item, they could not go back to previous questions, both for the CDR and the Ch-ICAR. All items required a response. Participants were explicitly told not to ask content-related questions, nor could they use supplementary materials (e.g., calculator, pen and paper). Most participants finished the Ch-ICAR within 25 minutes[10].

Data on school achievement were communicated to us at the end of the school year (June 2022) by the school principal.

Analyses

Analyses were conducted in SPSS (version 29.0; IMB Corp., 2022) and R (version 4.3.3; R Core Team, 2024). All R analysis codes as well as the variance-covariance matrix are available at: https://osf.io/6cwfs/?view_only=4a80feb83bd94c378d3bd06853820563.

To examine the internal structure of the test and the psychometric properties of individual items, we conducted a hierarchical (aligned with CHC-theory) Item Response Theory analysis (IRT; Chalmers, 2012), using the mirt package in R (version 1.42; Chalmers, 2023). The expected structure based on CHC-theory consists of two levels: several broad cognitive abilities at the first, lowest level (represented by the four Ch-ICAR subtests) and one general[11] cognitive ability at the second, highest level (Flanagan & McDonough, 2018). Items with standardized slopes below 0.20 were removed one at a time, starting with the item with the lowest loading. After each removal, the model was re-estimated. Subsequently, we conducted a Differential Item Functioning (DIF) analysis on the fitted hierarchical IRT model, using score-based structural change tests (Merkle et al., 2014), as implemented in the scDIFtest-package in R (version 0.1.1; Debeer, 2020). Items showing significant DIF, with a p-value below .05 after False Discovery Rate (FDR; Benjamini & Hochberg, 1995)

[10]Our time registration records encompassed the entire test administration, including the CDR-test and inter-test breaks, rather than specific segments. As a result, we can only provide an approximation of the time needed for completing the Ch-ICAR.

[11]When cognitive ability is explicitly linked to the theoretical framework of the CHC-model, we use the term "general cognitive ability" instead of “cognitive ability” (Flanagan & McDonough, 2018).

correction, were removed. The final model was re-estimated, and the model fit was evaluated.

To examine the distribution of scores, mean scores and standard deviations per subtest were calculated and a histogram of the Ch-ICAR total score was plotted. For this plot, the raw total scores were transformed into z-scores ($M = 0$, $SD = 1$). To examine potential sex differences in the total score, we conducted an independent samples t-test with sex as grouping variable and the Ch-ICAR total score as dependent variable.

The internal consistency of the different subtests and the Ch-ICAR as a whole is determined by calculating Cronbach's alpha and McDonald's omega[12] (Dunn et al., 2014; McDonald, 1999; Raykov, 1998) in R using the MBESS package (version 4.9.3; Kelley & Lai, 2012). To assess the reliability coefficients, we adhere to the COTAN guidelines for test quality evaluation (Evers et al., 2010). The COTAN rating system has established cut-off criteria for assessing reliability coefficients, which vary depending on the intended purpose of the test. COTAN distinguishes between three main purposes: tests for important decisions at individual level, tests for relatively less important decisions at individual level and tests for research at group level (Evers et al., 2010, p.34). Since our primary goal is to develop an instrument suitable for research purposes, we apply the cut-off criteria designated for research at group level. According to these criteria, reliability coefficients exceeding .70 are considered good, those ranging between .60 and .70 are deemed sufficient, while coefficients below .60 are deemed insufficient.

To examine the relationship with academic achievement, we calculated Pearson correlations and conducted mixed effects models for the different measures of academic achievement with the lmerTest package (version 3.1.3; Kuznetsova et al., 2017) in R. To

[12]When the assumption of tau-equivalence is violated (which is frequently the case in psychology), omega outperforms alpha (Dunn et al., 2014; Zinbarg et al., 2005). Omega has less risk of overestimation or underestimation of reliability, which makes it the preferred choice.

assess the unique contribution of the Ch-ICAR total score to the marginal explained variance in academic achievement, we contrasted the marginal explained variance between the full model and a model excluding the Ch-ICAR score (Nakagawa & Schielzeth, 2013; Shaw et al., 2023). The marginal explained variance was calculated via the MuMIn package (version 1.47.5; Bartoń, 2023).

Results

Internal structure and psychometric properties

In the initial hierarchical IRT model, with all items included, three items (MR5, MR8, and FA7) demonstrated low standardized slopes (i.e. <0.200), with values of 0.191, 0.199, and 0.110, respectively. We removed FA7, the item with the lowest slope, and re-estimated the model. In the second iteration, two items (MR5 and MR8) still had slopes below 0.200 (0.191 and 0.199, respectively), so MR5 was removed, and the model was estimated again. In the third iteration, MR8 alone exhibited a standardized slope below 0.200 (0.192) and was removed. By the fourth iteration, all remaining items had standardized slopes above the 0.200 threshold.

We conducted a DIF analysis on the fitted hierarchical IRT model (excluding items MR5, MR8 and FA7), which revealed significant DIF only in item NS6, with a Lagrange Multiplier (LM) statistic of 16.510 (df = 2, FDR-corrected $p = .007$). After removing item NS6, we re-estimated the hierarchical IRT model and evaluated the final model. Although the S-χ^2 statistic is not specifically designed for hierarchical IRT models, we used this statistic to assess item misfit, as the hierarchical structure of the model supports the use of summed item scores. The S-χ^2 test for three items (VR4, VR5, and NS4) yielded FDR-adjusted p-values below .05 (Table 3). However, after visually inspecting the observed response curves, these items were retained (Figure 1, 2, and 3). The S-χ^2 test results for all items are provided in Appendix 2D.

The standardized and unstandardized slopes for the final model are presented in Table 4, and the DIF statistics are shown in Table 5. The final model exhibited a good overall fit to the data: RMSEA: .027, 95% CI [.022, .031]; SRMR: .040; CFI: .976; TLI: .973[13]. The estimated regression weights for general cognitive ability on the four dimensions are as follows: Matrix Reasoning = 1.111, Verbal Reasoning = 1.390, Number Series = 1.845, and Figural Analogies = 0.654, indicating that general cognitive ability has a smaller load on the Figural Analogies dimension. Given the improved psychometric properties of this 27-item subset compared to the 31-item pilot version, we used this refined 27-item version for all further analyses.

Table 3

Item Fit Final Hierarchical IRT Model

Item	S_χ^2 Statistic	df	RMSEA	FDR adjusted p-value
VR4	28.1	12	.041	.048
VR5	35.0	14	.043	.020
NS4	35.9	10	.056	.002

[13]We use the cutoff values of Hu and Bentler (1999) to evaluate the goodness of fit: RMSEA <.06, SRMR <.08, and CFI & TLI > .95 indicate good fit.

Figure 1

Item Fit Plot Item VR4: Observed versus Expected Values

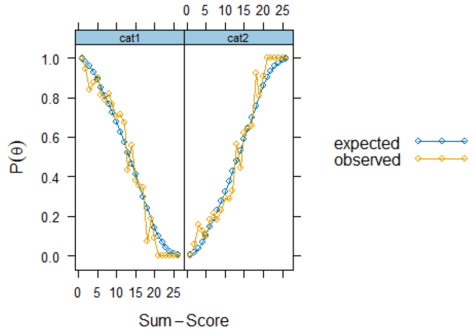


Figure 2

Item Fit Plot Item VR5: Observed versus Expected Values

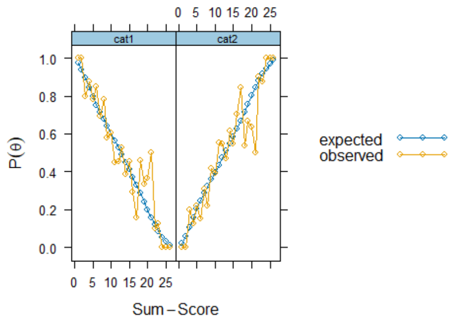


Figure 3

Item Fit Plot Item NS4: Observed versus Expected Values

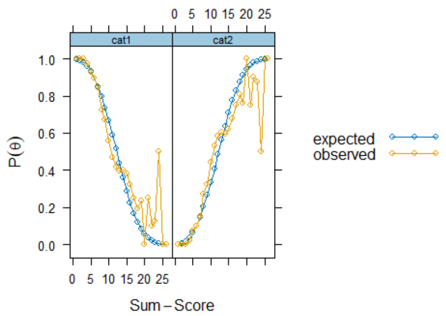


Table 4

Standardized and Unstandardized Slopes Final Hierarchical IRT Model

Item	Unstd. Intercept	Unstd. MR slope	Std. MR slope	Unstd. VR slope	Std. VR slope	Unstd. NS slope	Std. NS slope	Unstd. FA slope	Std. FA slope
MR1	-0.530	0.465	0.231	0.000	0.000	0.000	0.000	0.000	0.000
MR2	-0.030	1.120	0.444	0.000	0.000	0.000	0.000	0.000	0.000
MR3	-1.316	0.595	0.285	0.000	0.000	0.000	0.000	0.000	0.000
MR4	-1.082	0.849	0.373	0.000	0.000	0.000	0.000	0.000	0.000
MR6	-1.036	0.618	0.294	0.000	0.000	0.000	0.000	0.000	0.000
MR7	-1.726	0.456	0.227	0.000	0.000	0.000	0.000	0.000	0.000
VR1	1.742	0.000	0.000	0.589	0.229	0.000	0.000	0.000	0.000
VR2	1.312	0.000	0.000	1.009	0.344	0.000	0.000	0.000	0.000
VR3	-0.075	0.000	0.000	0.956	0.332	0.000	0.000	0.000	0.000
VR4	-0.724	0.000	0.000	1.392	0.414	0.000	0.000	0.000	0.000
VR5	-0.362	0.000	0.000	0.838	0.303	0.000	0.000	0.000	0.000
VR6	-1.578	0.000	0.000	0.665	0.253	0.000	0.000	0.000	0.000
VR7	-1.532	0.000	0.000	0.532	0.210	0.000	0.000	0.000	0.000
VR8	-1.531	0.000	0.000	1.058	0.355	0.000	0.000	0.000	0.000
NS1	1.827	0.000	0.000	0.000	0.000	0.884	0.242	0.000	0.000
NS2	0.058	0.000	0.000	0.000	0.000	1.106	0.283	0.000	0.000
NS3	0.349	0.000	0.000	0.000	0.000	1.729	0.360	0.000	0.000
NS4	-0.671	0.000	0.000	0.000	0.000	2.274	0.397	0.000	0.000
NS5	-0.428	0.000	0.000	0.000	0.000	3.067	0.428	0.000	0.000
NS7	-1.572	0.000	0.000	0.000	0.000	1.331	0.316	0.000	0.000
NS8	-2.067	0.000	0.000	0.000	0.000	1.214	0.300	0.000	0.000
FA1	-0.370	0.000	0.000	0.000	0.000	0.000	0.000	1.651	0.729
FA2	-0.075	0.000	0.000	0.000	0.000	0.000	0.000	1.919	0.753
FA3	-0.164	0.000	0.000	0.000	0.000	0.000	0.000	2.344	0.778
FA4	0.174	0.000	0.000	0.000	0.000	0.000	0.000	1.218	0.665
FA5	-2.665	0.000	0.000	0.000	0.000	0.000	0.000	0.378	0.315
FA6	-1.756	0.000	0.000	0.000	0.000	0.000	0.000	0.626	0.467

Note. VR = Verbal Reasoning, MR = Matrix Reasoning, NS = Number Series, FA = Figural Analogies.

Table 5

Item-Wise Score-Based DIF Detection Final Hierarchical IRT Model

item	item type	number of est. parameters	LM test statistic	p-value	FDR adjusted p-value
MR1	2PL	2	1.962	0.375	0.562
MR2	2PL	2	2.401	0.301	0.562
MR3	2PL	2	2.256	0.324	0.562
MR4	2PL	2	1.268	0.531	0.677
MR6	2PL	2	5.239	0.073	0.328
MR7	2PL	2	0.379	0.827	0.894
VR1	2PL	2	0.563	0.755	0.849
VR2	2PL	2	2.771	0.250	0.562
VR3	2PL	2	0.235	0.889	0.896
VR4	2PL	2	7.652	0.022	0.147
VR5	2PL	2	1.101	0.577	0.677
VR6	2PL	2	3.157	0.206	0.506
VR7	2PL	2	2.031	0.362	0.562
VR8	2PL	2	10.624	0.005	0.069
NS1	2PL	2	3.911	0.142	0.382
NS2	2PL	2	5.808	0.055	0.296
NS3	2PL	2	4.216	0.121	0.382
NS4	2PL	2	10.047	0.007	0.069
NS5	2PL	2	2.090	0.352	0.562
NS7	2PL	2	9.731	0.008	0.069
NS8	2PL	2	4.370	0.112	0.382
FA1	2PL	2	1.515	0.469	0.666
FA2	2PL	2	4.027	0.134	0.382
FA3	2PL	2	0.219	0.896	0.896
FA4	2PL	2	1.105	0.575	0.677
FA5	2PL	2	1.124	0.570	0.677
FA6	2PL	2	2.144	0.342	0.562

Internal consistency

The full Ch-ICAR showed good internal consistency ($\alpha = .82$; $\omega = .82$, 95% CI_{ω} [.81, .84]). At subtest level, two of the four subtests also showed good internal consistency: Number Series ($\alpha = .73$; $\omega = .75$, 95% CI_{ω} [.72, .77]) and Figural Analogies ($\alpha = .77$; $\omega = .81$, 95% CI_ω [.79, .83]). The internal consistency of the other two subtests was less favorable: Verbal Reasoning had $\alpha = .60$; $\omega = .61$, 95% CI_ω [.57, .65] and Matrix Reasoning had $\alpha = .48$; $\omega = .48$, 95% CI_ω [.42, .54].

Distribution of Ch-ICAR scores

Table 6 presents the means and standard deviations of the Ch-ICAR (subtest) scores, with separate values provided for boys and girls. Notably, the maximum scores are 8 for Verbal Reasoning, 6 for both Matrix Reasoning and Figural Analogies, 7 for Number Series, and 27 for the full Ch-ICAR. Results of the independent samples t-test showed no significant difference in Ch-ICAR total score between males (M = 10.91, SD = 5.20) and females (M = 10.54, SD = 4.99): $t(812) = 1.02$, $p = .31$, *Cohen's d* = .07. The histogram of the standardized Ch-ICAR total scores is depicted in Figure 4.

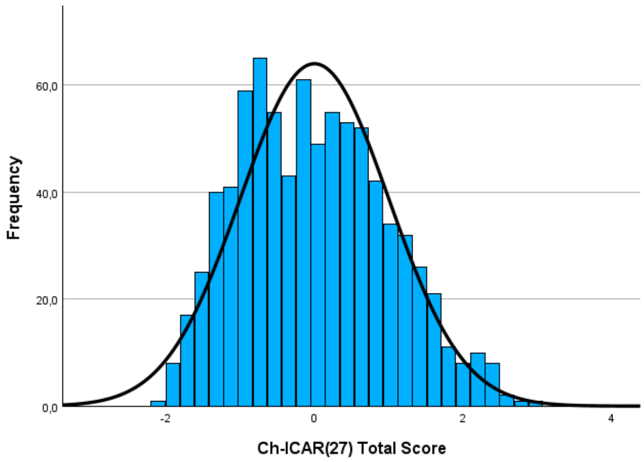
Table 6

Means and Standard Deviations of the Ch-ICAR (Subtest) Scores

(sub)test	M (*SD*)	M *(SD)* Boys	M *(SD)* Girls
Verbal Reasoning	3.49 *(1.81)*	3.64 *(1.86)*	3.28 *(1.74)*
Matrix Reasoning	1.86 *(1.42)*	1.79 *(1.45)*	1.94 *(1.39)*
Number Series	3.19 *(1.99)*	3.41 *(2.00)*	2.93 *(1.95)*
Figural Analogies	2.22 *(1.85)*	2.07 *(1.81)*	2.40 *(1.88)*
Full Ch-ICAR	10.76 *(5.11)*	10.91 *(5.20)*	10.54 *(4.99)*

Figure 4

Histogram of Standardized Ch-ICAR Total Scores



Relation with academic performance

All measures of academic performance were positively and significantly correlated with both the subtests and the full Ch-ICAR (Table 7). Correlations with CDR were consistently stronger than those with the other measures of academic performance. However, given that two out of four subtests demonstrate limited reliability, we recommend caution when interpreting individual subtest scores independently.

Table 7

Correlations Between Academic Performance and the Ch-ICAR (Sub)tests

Academic performance	Verbal Reasoning	Matrix Reasoning	Number Series	Figural Analogies	Full Ch-ICAR
GPA	.35**	.27**	.29**	.28**	.42**
Dutch	.33**	.23**	.28**	.25**	.38**
Mathematics	.40**	.29**	.38**	.32**	.49**
Natural sciences	.34**	.23**	.30**	.25**	.40**
Technics	.20**	.21**	.14**	.20**	.26**
CDR	.58**	.31**	.54**	.34**	.63**

Note. ** $p < .01$.

Subsequently, we conducted three mixed effects models, each with a different dependent variable: GPA, mathematics score and CDR score. In each model, we included the Ch-ICAR total score (continuous), sex (categorical) and all four SES indicators[14] (all categorical) as fixed effects. Additionally, school was included as a random factor in each model to account for the multilevel nature of the data. To address potential multicollinearity concerns, we assessed the variance inflation factor (VIF) for all predictors. All VIF values ranged from 1.01 to 1.30, which is well below the commonly cited thresholds[15] for problematic multicollinearity in the model (Marcoulides & Raykov, 2019; O'Brien, 2007). Table 8 includes the results of the mixed effects models. The models achieved a marginal R^2 of .32, .34 and .40 for GPA, mathematics score and CDR score as dependent variable respectively, indicating that up to 40% of the variance in academic performance is explained by the fixed effects portion of the model (Nakagawa & Schielzeth, 2013; Shaw et al., 2023). The Ch-ICAR total score uniquely explained 18%, 26% and 35% of the within school variance in GPA, math score and CDR score respectively via its fixed effects.

[14]All SES indicators were significantly correlated with GPA, math score and CDR score, with the exception of *Educational Level Parent 1*, which showed no correlation with GPA and math score. See Appendix 2E for the full correlation table.

[15]"Not uncommonly a VIF of 10 or even one as low as 4 have been used as rules of thumb to indicate excessive or serious multicollinearity" (O'Brien, 2007, p.674).

Table 8

Estimates, Standard Errors, t-values and p-values for the Fixed and Random Effects of the Three Linear Mixed Effect Models

	GPA				Mathematics Score				CDR Score			
	β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*
Fixed Effects												
Intercept	-.99	.11	-8.94	<.001	-1.07	.11	-9.42	<.001	-1.27	.09	-13.73	<.001
Ch-ICAR	.08	.01	11.94	<.001	.10	.01	14.40	<.001	.12	.01	18.38	<.001
Sex	.50	.07	7.03	<.001	.14	.07	1.98	.05	.18	.06	2.83	<.01
Language	-.06	.14	-.40	.69	-.13	.14	-.90	.37	-.24	.13	-1.94	.05
Bursary	-.47	.08	-5.68	<.001	-.45	.08	-5.39	<.001	-.06	.07	-.80	.42
Education Parent 1	.18	.14	1.25	.21	.22	.14	1.56	.12	-.10	.12	-.84	.40
Education Parent 2	-.31	.15	-2.08	.04	-.15	.15	-1.04	.30	-.24	.12	-2.03	.04
Random Effects	Variance	SD			Variance	SD			Variance	SD		
School												
(Intercept)	.04	.19			.05	.21			.01	.10		

Note. Reference level for Sex = boys; reference level for Language = Dutch as home language; reference level for Bursary = no bursary; reference level for Education parent 1 and 2 = highest attained degree is secondary education or higher.

Discussion

In Study 1, we aimed to evaluate the psychometric properties of the Ch-ICAR by examining four key aspects of the test. Results from the final hierarchical IRT analysis, excluding four items (MR5, MR8, NS6, and FA7), provided support for the Ch-ICAR's psychometric robustness. All remaining items demonstrated adequate discriminatory power, and no DIF was detected. Furthermore, the good model fit aligns with expectations based on CHC theory, which posits several broad cognitive abilities at the first, lower level and general cognitive ability at the second, higher level (Flanagan & McDonough, 2018). We thus utilized the 27-item subset in all subsequent analyses and recommend that future research also adopt this subset when using the Ch-ICAR.

Regarding internal consistency, analyses revealed that the full Ch-ICAR shows good internal consistency, which contributes to the quality of the test. At the subtest level, however, reliability estimates are mixed: While Number Series and Figural Analogies demonstrate good internal consistency, Matrix Reasoning and Verbal Reasoning exhibit insufficient and marginally sufficient reliability respectively. These findings align with prior research on the original ICAR16, where similar[16] reliability challenges were reported for Matrix Reasoning ($\omega = .55$) and Verbal Reasoning ($\omega = .61$; Condon & Revelle, 2014; Young et al., 2019). The consistent pattern of low reliabilities across studies suggests that these two subtests may be capturing multiple underlying factors rather than a single dimension. Given that two out of four subtests demonstrate limited reliability, we recommend caution when interpreting individual subtest scores independently. However, reliability and validity are always a function of both the instrument and the specific population being tested (Bannigan & Watson,

[16]Of course, reliability is influenced by the length of a test, so comparisons between tests of differing lengths should be made with careful consideration and nuance.

2009), so future research is essential to assess whether these subtests might perform more reliably in other populations.

With respect to the distribution of the Ch-ICAR scores, our analysis showed no evidence of sex bias. This result aligns with previous research, which suggests that although small sex differences in specific cognitive skills may exist, overall cognitive ability is generally equal across sexes (e.g., Deary et al., 2007; Fergusson & Horwood, 1997; Nisbett et al., 2012).

Regarding the relation with academic performance, we found that cognitive ability as measured by the Ch-ICAR uniquely explained about 18% of the within school variance in GPA, 26% of the within school variance in the score for the school subject mathematics, and 35% of the within school variance in the score on the CDR. The proportion of explained variance in school achievement is somewhat lower than what is commonly reported for educational tests (i.e. medium to high, depending on the way academic performance is measured and analyzed, see for instance Kort et al., 2002; Deary et al., 2007; Koenig et al., 2008), but acceptable considering the rather short nature of the Ch-ICAR. Moreover, it is crucial to highlight that we have controlled for SES and sex, and accounted for the nested structure of our data, enhancing the robustness of the results. Furthermore, the correlations in our study are similar to the magnitude of the correlations with academic achievement found in the initial validation study of the ICAR16 (Condon & Revelle, 2014), which confirms the present set as a useful contribution to the ICAR project.

In conclusion, the Ch-ICAR seems promising as a quick, cost-efficient (i.e., free to use) and useful tool for researchers to obtain a reliable and valid, sex-neutral assessment of cognitive ability.

Study 2: Cross-validation of the Children's ICAR

Study 2 sought to cross-validate the Ch-ICAR with two established intelligence tests: the Raven's 2 Progressive Matrices (RPM; McLeod & McCrimmon, 2021) and the CoVaT-CHC Basic Version (CoVaT; Magez et al., 2015). We examined the relation between the Ch-ICAR and cognitive ability, estimated by RPM. Additionally, we examined the relations between the Ch-ICAR subtests and three broad cognitive abilities estimated by the CoVaT: Fluid intelligence (Gf), Visuospatial processing (Gv), and Crystallized intelligence (Gc), as well as the relation between the Ch-ICAR and the CoVaT total score. A priori we expected significant correlations of the Ch-ICAR with on the one hand RPM and on the other CoVaT.

As mentioned in the general introduction, the ICAR project does not start from an a priori theoretical framework of cognitive ability. Only one study explored the connection between the original four ICAR-(sub)tests and the CHC-model (Young & Keith, 2020). Results showed a strong correlation between the ICAR16 total score and general cognitive ability, estimated by the WAIS-IV. At subtest level, Young and Keith (2020) found that Letter and Number Series correlated most strongly with tests of Gf, whereas the remaining three subtests (Verbal Reasoning, Three-dimensional Rotation, and Matrix Reasoning) showed the strongest correlations with tests of Gv. Regarding Verbal Reasoning, Young and Keith (2020) themselves expressed surprise at the results given the inconsistency with the CHC-theory. They considered it highly plausible that the results stemmed from a statistical artifact or the small sample size, and hence pleaded for replications. Regarding Matrix Reasoning, the authors suggested that it is likely that it calls on both Gf and Gv but recommended caution in interpretation given the low internal consistency.

Building on cognitive ability theory, we anticipated that the Ch-ICAR subtests Number Series and Matrix Reasoning mainly tap into Gf; that Figural Analogies draws on both Gf and Gv; and that Verbal Reasoning calls on Gf and Gc, since this subtest requires both mathematical operations and general knowledge. However, we did not rule out the possibility that the results regarding Verbal Reasoning and Matrix Reasoning are in line with the earlier research of Young and Keith (2020), and primarily assess Gv. It is important to note here that the Ch-ICAR does not aspire to measure IQ like the Wechsler scales. Instead, the Ch-ICAR strives to be a concise, open-source measure of cognitive ability for research purposes, in line with the overarching aim of the ICAR project (Condon & Revelle, 2014).

Method

Participants

Data for cross-validation with RPM were collected from 91 pupils (N_{RPM} = 91, male 40%), who were in their second year of secondary education (M_{age} = 12.99, SD_{age} = .57). In this sample, 3% of the participants had another home language than Dutch. Data for cross-validation with the CoVaT were collected from 96 pupils (N_{CoVaT} = 96, male 50%), who were in their last year of primary education (M_{age} = 11.05, SD_{age} = .27). In this sample, 3% of the participants had another home language than Dutch. Table 2 offers a comprehensive overview of the demographics and associated statistics for each sample.

Materials

Children's ICAR. All pupils completed the Ch-ICAR (see Study 1 for a detailed description of the instrument). Based on recommendations from Study 1, we utilized the 27-item subset for all analyses. The means and standard deviations of the Ch-ICAR (subtest) scores for the CoVaT and RPM samples are presented in Tables 9 and 10 respectively, with separate values reported for boys and girls[17]. The full Ch-ICAR showed good internal

---

[17]In Flanders, primary education is comprehensive, with no ability grouping and open access, while secondary education is specialized and ability-based (Seghers et al., 2019). Study 2 focused on examining the Ch-ICAR's correlation with other cognitive ability tests, rather than sampling a representative population. Consequently, sex balance across ability groups in the RPM sample was not specifically monitored, as this was not a focus of the study. Nevertheless, independent samples t-tests showed no significant difference in Ch-ICAR total scores between males and females in either sample: CoVaT sample: $t(94) = .82$, $p = .41$, *Cohen's d* = .17; RPM

consistency in both samples, with McDonald's omega values of .81 in the RPM sample and .84 in the CoVaT sample. At subtest level, internal consistency was sufficient to good for three of the four subtests (Number Series, Figural Analogies and Verbal Reasoning) in both samples, with McDonald's omega values ranging from .60 to .77. However, internal consistency for the Matrix Reasoning subtest was insufficient (ω = .31 in the RPM sample; ω = .52 in the CoVaT sample). Comprehensive internal consistency statistics, including both McDonald's omega and Cronbach's alpha, are provided in Appendix 2F.

Table 9

Means and Standard Deviations of the Ch-ICAR (Subtest) Scores for CoVaT Sample

(sub)test	M (*SD*)	M (*SD*) Boys	M *(SD)* Girls
Verbal Reasoning	3.93 *(1.78)*	4.17 *(2.01)*	3.69 *(1.49)*
Matrix Reasoning	2.55 *(1.52)*	2.46 *(1.52)*	2.65 *(1.54)*
Number Series	3.58 *(1.98)*	4.13 *(1.94)*	3.04 *(1.88)*
Figural Analogies	2.22 *(1.80)*	1.98 *(1.76)*	2.46 *(1.82)*
Full Ch-ICAR	12.28 *(5.32)*	12.73 *(5.53)*	11.83 *(5.12)*

Note. The maximum scores are 8 for Verbal Reasoning, 6 for both Matrix Reasoning and Figural Analogies, 7 for Number Series, and 27 for the full Ch-ICAR.

Table 10

Means and Standard Deviations of the Ch-ICAR (Subtest) Scores for RPM Sample

(sub)test	M (*SD*)	M (*SD*) Boys	M *(SD)* Girls
Verbal Reasoning	3.98 *(1.99)*	3.61 *(1.78)*	4.22 *(2.10)*
Matrix Reasoning	2.51 *(1.34)*	2.42 *(1.50)*	2.56 *(1.23)*
Number Series	4.12 *(2.06)*	4.11 *(2.05)*	4.13 *(2.08)*

sample: $t(89) = -1.48$, $p = .14$, *Cohen's d* = -.32. Furthermore, no DIF was detected in any items in the refined hierarchical IRT model (excluding items MR5, MR8, NS6 and FA7).

Figural Analogies	2.98 *(1.71)*	2.47 *(1.87)*	3.31 *(1.53)*
Full Ch-ICAR	13.58 *(5.09)*	12.61 *(5.40)*	14.22 *(4.83)*

Note. The maximum scores are 8 for Verbal Reasoning, 6 for both Matrix Reasoning and Figural Analogies, 7 for Number Series, and 27 for the full Ch-ICAR.

Raven's 2 Progressive Matrices. We used the Raven's 2 Digital Short Form (McLeod & McCrimmon, 2021). The test starts with three example items followed by three practice items. The actual test consists of 24 items and has a time limit of 20 minutes. Each participant received a unique but comparable subset of items, automatically generated from the Raven's 2 item bank. The marginal reliability coefficient of the Raven's 2 Digital Short Form, based on American data, is .80 (Dimitrov, 2003). Per participant, a score report was obtained via Q-Global (Pearson, 2023). The report describes among others: a Total Raw Score (TRS), a Skill Score (SkS) and a Scaled Score (ScS). For our analyses we used the ScS[18]. In our sample, the mean ScS was $M = 96.44$, $SD = 12.18$.

CoVaT-CHC Basic Version. The full CoVaT (Magez et al., 2015) encompasses five broad cognitive abilities (BCA): Fluid intelligence (Gf), Visuospatial processing (Gv), Crystallized intelligence (Gc), Short-Term Memory (Gsm), and Processing Speed (Gs). Each BCA is measured by two subtests, except Gs, which is measured by only one. The CoVaT offers strong psychometric properties (Magez, 2019) and has received the highest quality label of the Belgian Federation for Psychologists (BFP, 2020). The complete CoVaT administration takes two and a half to three hours, but individual administration may be shorter. Due to time constraints, we only administered the six subtests measuring Gf, Gv, and Gc.

Gf is assessed by the subtests Point Sequences and Figure Sequences. Point Sequences

[18]The Skill Score converts each participant's test performance to a common and equal interval scale, regardless of test version or differences between item sets (Pearson, 2023). Skill Scores can be directly compared with each other. The Scaled Score is based on the Skill Score and the participant's age. The Scaled Score is a standardized score with a mean of 100 and a standard deviation of 15, and can be compared to the score on other intelligence tests.

comprises 15 items where participants have to draw points to complete a sequence. Figure Sequences comprises 25 items and requires participants to draw figures to complete a sequence. Gc is measured via the subtests Shifts and Opposites, each consisting of 35 items. In Shifts, participants have to select the word that does not fit in a series, while Opposites requires participants to select the opposite word from a target word. Gv is evaluated using the subtests Rotated Figures (20 items) and Folding Boxes (26 items), assessing participants' ability to identify equal, albeit rotated, figures and match folded with unfolded boxes, respectively. Each subtest has a specified time limit outlined in the manual. The reliability coefficients of Gf, Gv, and Gc are .96, .91 and .87 respectively[19]. The means and standard deviations of the BCAs in our sample are shown in Table 11. We used the pen-and-paper version and administered the instrument in group.

Table 11

Means and Standard Deviations of the CoVaT (BCAs)

CoVaT (BCA)	*M* (*SD*)
Gv	54.76 *(17.59)*
Gf	67.32 *(14.97)*
Gc	54.42 *(9.86)*
CoVaT weighted overall score[20]	298.24 *(54.84)*

Procedure

For cross-validation with RPM, participants completed RPM and the Ch-ICAR on the same day. For cross-validation with the CoVaT, participants completed the CoVaT and the Ch-ICAR on different days, with minimum 7 and maximum 35 days between both tests (M_{days}

[19]The reliability of the BCA indices was estimated based on the formula of Lienert (Stinissen et al., 1975), which uses the observed standard deviations (σ) per subtest and the observed intercorrelations between subtests.
[20]The weighted CoVaT overall score = 2*Gf + 2*Gc + Gv.

= 16.33, SD_{days} = 16.17). We introduced a time delay specifically for the Ch-ICAR-CoVaT test-taking to prevent pupils from experiencing cognitive overload before beginning the subsequent test, given that the CoVaT spanned about three hours. Since the Raven only lasted 20 minutes, no time delay was deemed necessary between the Raven and the Ch-ICAR. The order of testing alternated between classes, in both samples.

Analyses

Analyses were conducted in SPSS (version 29.0; IMB Corp., 2022) and R (version 4.3.3; R Core Team, 2024). All R analysis codes as well as the variance-covariance matrix are available at: https://osf.io/6cwfs/?view_only=4a80feb83bd94c378d3bd06853820563.

To examine the relationship between the Ch-ICAR and cognitive ability, we calculated Pearson correlations between the observed Ch-ICAR (subtest) scores and the Scaled Score derived from the RPM, as well as between the observed Ch-ICAR (subtest) scores and the CoVaT's weighted overall score, and its measures of the broad cognitive abilities Gf, Gv, and Gc. In line with Young and Keith's (2020) cross-validation research, we also tested a final integrated CFA model informed by the theoretical expected relationships. The CFA was conducted using the lavaan package (version 0.6.17; Rosseel, 2012), with maximum likelihood (ML) as estimator and the Satorra-Bentler correction for non-normality. The model allowed for correlations between the broad cognitive abilities to account for the positive manifold (i.e., the psychometric phenomenon that all (sub)tests that measure facets of cognitive ability correlate; Burgoyne et al., 2022), and for correlated errors for the Ch-ICAR subtests to account for potential method covariance (Young & Keith, 2020).

Results

Cross-validation with the RPM

The correlations between the Ch-ICAR subtests and the ScS range from r = .39 to r = .59 (Table 12). The Ch-ICAR total score showed the largest correlation with the ScS: r = .67.

Table 12

Correlations Between the Ch-ICAR and RPM

(sub)test	1	2	3	4	5	6
1. RPM: ScS	1					
2. Ch-ICAR: VR	.59**	1				
3. Ch-ICAR: MR	.43**	.46**	1			
4. Ch-ICAR: NS	.48**	.47**	.33**	1		
5. Ch-ICAR: FA	.39**	.30**	.34**	.19	1	
6. Full Ch-ICAR	.67**	.80**	.69**	.74**	.62**	1

Note. ScS = Scaled Score, VR = Verbal Reasoning, MR = Matrix Reasoning, NS = Number Series, FA = Figural Analogies.
$^{**}p < .01$.

Cross-validation with the CoVaT

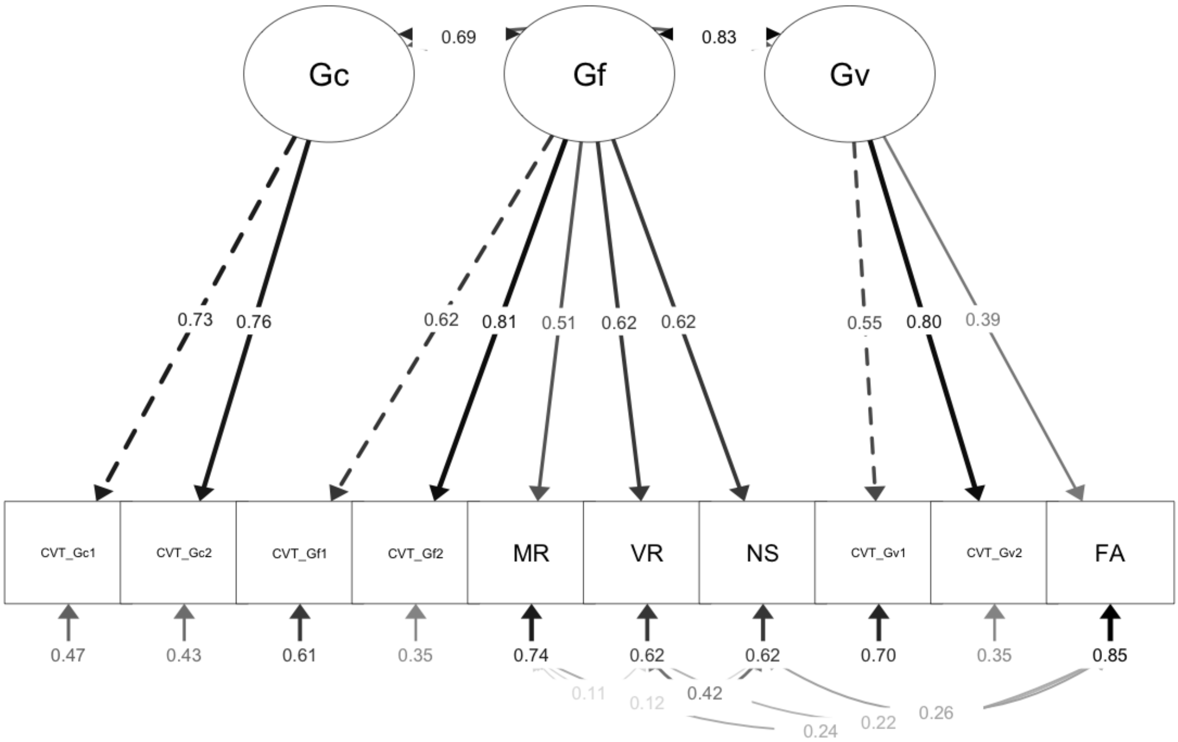
The correlations between the Ch-ICAR and CoVaT scores are depicted in Table 13. Number Series, Verbal Reasoning and Matrix Reasoning showed the strongest correlation with Gf, while Figural Analogies correlated most strongly with Gv, when considering all BCAs. However, given that two out of four subtests demonstrate limited reliability, we recommend caution when interpreting individual subtest scores independently.

The final integrated CFA model (Figure 5) showed a moderate fit for the data: χ^2 (26): 41.305; RMSEA = .081, 90% C.I. [.026, .125]; CFI = .944, TLI = .904, SRMR = .060. Given that we a priori expected Verbal Reasoning and Figural Analogies to show equally high correlations with Gf and Gc; and Gf and Gv respectively, we initially allowed those two subtests to crossload on both. However, compared to the final integrated model, crossloading Verbal Reasoning did not significantly improve the model fit: $\chi^2(25)$: 39.574; RMSEA = .081, 90% C.I. [.024, .126]; CFI = .947, TLI = .904, SRMR = .058; $\Delta\chi^2(1) = 1.774$, $p = .183$. Similarly, crossloading Figural Analogies also had no significant impact: χ^2 (25): 40.667;

RMSEA = .084, 90% C.I. [.030, .129]; CFI = .942, TLI = .896, SRMR = .061, $\Delta\chi^2(1) = .085$, $p = .771$. Therefore, we opted for the more parsimonious model without crossloadings.

Figure 5

Final Integrated CFA Model



Note. CVT_... = CoVaT subtest, VR = Verbal Reasoning scale score, MR = Matrix Reasoning scale score, NS = Number Series scale score, FA = Figural Analogies scale score.

Table 13

Correlations Between the Ch-ICAR (Sub)tests, the CoVaT Overall Score and Three Broad Cognitive Abilities Estimated by the CoVaT

(sub)test	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. CoVaT weighted overall score	1								
2. CoVaT: Gv	.79**	1							
3. CoVaT: Gf	.90**	.63**	1						
4. CoVaT: Gc	.71**	.35**	.43**	1					
5. Ch-ICAR: VR	.54**	.28**	.53**	.45**	1				
6. Ch-ICAR: MR	.47**	.38**	.40**	.37**	.40**	1			
7. Ch-ICAR: NS	.55**	.35**	.51**	.44**	.62**	.39**	1		
8. Ch-ICAR: FA	.31**	.29**	.26*	.21*	.34**	.35**	.40**	1	
9. Full Ch-ICAR	.62**	.43**	.57**	.49**	.79**	.68**	.82**	.70**	1

Note. VR = Verbal Reasoning, MR = Matrix Reasoning, NS = Number Series, FA = Figural Analogies.
*p < .05, **p < .01.

Discussion

In Study 2, we cross-validated the Ch-ICAR with the RPM in a sample of 91 pupils in secondary education; and with the CoVaT in a sample of 96 pupils in primary education. Analyses revealed high correlations between the Ch-ICAR total score and cognitive ability, estimated by RPM and CoVaT. The magnitude of the relationships is in line with prior cross-validation research with the ICAR16 (Young & Keith, 2020). As expected, the Ch-ICAR is mainly a measure of Gf, making it a viable measure for non-verbal cognitive ability, specifically.

At subtest level, the correlation matrix in combination with the integrated CFA model suggests that the subtests tend to measure what we theoretically expected. We found Verbal Reasoning to predominantly assess Gf and, to a somewhat lesser extent, Gc. These results are consistent with the theory and empower the statement of Young and Keith (2020) that their findings (Verbal Reasoning mainly tapping into Gv) are presumably a statistical artifact or a product of the small sample.

Regarding Figural Analogies, it is notable that this subtest shows rather small correlations with the CoVaT across the board. When we take a closer look, several reasons emerge as to why this subtest does not function as expected. First, the scores on this subtest are not normally distributed at all (Appendix 2G); many participants either get four questions right or none at all. Secondly, and related to the first reason, there appears to be an issue with increasing difficulty: While about half of the participants answered items 1-4 correctly, the correct response rate dropped to 11% for item 5 and 20% for item 6 (Appendix 2H displays the proportion of correct responses for each item across all samples). Additionally, as we aimed for standardization in the testing procedure, we presented the subtests in a fixed sequence, with Figural Analogies as the final subtest. While this procedure adheres to common practice (Flanagan & Alfonso, 2017; Weiss et al., 2019), it introduces the potential for fatigue effects to impact test performance. To address these issues, we recommend conducting further research to assess the validity of Figural Analogies with other items of the subtest that vary in difficulty.

The subtest Matrix Reasoning correlates most strongly with Gf, as expected; however, its low reliability requires careful interpretation. Future research should explore ways to enhance the quality of Matrix Reasoning, as its validity issues reflect broader challenges within the ICAR project (Condon & Revelle, 2014, p. 57; Young et al., 2019, p. 2). For the Ch-ICAR specifically, developing a larger set of easier items and incorporating simple, non-verbal instructions may enhance the subtest's internal consistency and further improve the overall quality of the Ch-ICAR. Additionally, one could consider replacing Matrix Reasoning items with alternative item types, such as Progressive Matrices items, as those use a similar task paradigm but offer much stronger reliability.

In conclusion, the Ch-ICAR total score exhibits a strong correlation with cognitive ability, establishing itself as a valid and freely available measure for assessing children's cognitive abilities in research settings. Nonetheless, at this stage, we advise against using and interpreting subtests independently.

As the sample sizes of Study 2 are rather small, the current results should be interpreted with caution and replications in larger samples are needed. Nevertheless, the results of both cross-validations yield the same conclusions, which removes part of the potential bias risk.

General conclusion

In developing the Ch-ICAR, our aim was to expand the scope of the ICAR project, considering that its suitability for children had not been previously assessed (Dworak et al., 2021; The International Cognitive Ability Resource Team, 2014). Additionally, we sought to address the dearth of psychometrically robust cognitive ability measures that are freely accessible, brief in administration time, and tailored to children.

The results of Study 1 and 2 provide support for the utility of the full Ch-ICAR as a measurement for children's cognitive abilities within a research context. This was demonstrated by the observed correlations between the Ch-ICAR and the RPM and CoVaT, two established intelligence tests (McLeod & McCrimmon, 2021; Magez et al., 2015) and by the established relationship with academic performance. The magnitude of the correlation coefficients is in line with prior validation research of the ICAR16 (Condon & Revelle, 2014; Young & Keith, 2020).

At this point, however, we advise against the use of the Ch-ICAR subtests in isolation, in contrast with the use of the battery as a whole. Further research should ascertain whether the subtests can reliably stand alone and yield meaningful interpretations. Therefore, replications of the relations with the broad cognitive abilities are necessary, as well as improvements in the subtests Matrix Reasoning and Figural Analogies.

The primary constraint of this investigation pertains to the restricted scope of the Ch-ICAR, which solely assessed a limited subset of ICAR items in children. Consequently, the possibility cannot be dismissed that an alternative or more extensive item selection may yield superior performance as a child-adapted version of the ICAR. Another consideration involves the absence of established norms for the Ch-ICAR. Given the dynamic developmental changes in cognitive processing efficiency during childhood and adolescence (Flanagan & McDonough, 2018), it is imperative to consider factors such as age or grade level when interpreting and comparing Ch-ICAR results. Therefore, we underscore the significance of future research endeavors aimed at establishing age- or grade-based norms to ensure precise interpretation and comparison of Ch-ICAR outcomes.

References

Ali, A., Ambler, G., Strydom, A., Rai, D., Cooper, C., McManus, S., … Hassiotis, A. (2013). The relationship between happiness and intelligent quotient: The contribution of socio-economic and clinical factors. *Psychological Medicine, 43,* 1303-1312. https://doi.org/10.1017/S0033291712002139

Aubry, A., & Bourdin, B. (2018). Short forms of Wechsler scales assessing the intellectually gifted children using simulation data. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00830

Bannigan, K., & Watson, R. (2009). Reliability and validity in a nutshell. *Journal of Clinical Nursing, 18,* 3237-3243. https://doi.org/ 10.1111/j.1365-2702.2009.02939.x

Bartoń, K. (2023). *MuMIn: Multi-Model Inference* (Version 1.47.5) [Software]. Retrieved June 3, 2024, from https://cran.r-project.org/web/packages/MuMIn/index.html

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology, 57*, 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

BFP Testcommissie (2020). Testfiche CoVaT-CHC Basisversie. BFP. Retrieved December 13, 2023, from https://www.bfp-fbp.be/_files/ugd/8d37d9_8647ceeb137449b5b967fbf48fa89f9e.pdf

Burgoyne, A. P., Mashburn, C. A., Tsukahara, J. S., & Engle, R. W. (2022). Attention control and process overlap theory: Searching for cognitive processes underpinning the positive manifold. *Intelligence, 91,* 101629. https://doi.org/10.1016/j.intell.2022.101629

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software, 48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Chalmers, R. P. (2023). *mirt: Multidimensional item response theory* (Version 1.35) [R package]. Retrieved November 14, 2024, from https://cran.r-project.org/web/packages/mirt/mirt.pdf

Condon, D., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21. https://doi.org/10.1016/j.intell.2006.02.001

Debeer, D. (2020). *scDIFtest: Item-Wise Score-Based DIF Detection* (Version 0.1.1) [R package]. Retrieved November 14, 2024, from https://cran.r-project.org/web/packages/scDIFtest/scDIFtest.pdf

Desoete, A., & Roeyers, H. (2006). *Cognitieve Deelhandelingen van het Rekenen (CDR). Handleiding & testprotocol [Cognitive Developmental skills in Arithmetics. Manual & testprotocol].* Herenthals: VVL.

Dimitrov, D. M. (2003). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement, 6,* 440-458. https://doi.org/10.1177/0146621603258786

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology, 105,* 399-412. https://doi.org/10.1111/bjop.12046

Dworak, E., Revelle, W., Doebler, P., & Condon, D. (2021). Using the International Cognitive Ability Resource as an open source tool to explore individual differences in cognitive ability. *Personality and Individual Differences*, *169*, Article 109906. https://doi.org/10.1016/j.paid.2020.109906

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN review system for evaluating test quality.* Boom Uitgevers Amsterdam. Retrieved June 3, 2024, from https://psynip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf

Evers, A., Muniz, J., Bartram, D., Boben, D., Egeland, J., Fernandez-Hermida, J. R., … Urbanek, T. (2012). Testing practices in the 21st century Developments and European psychologists' opinions. *European Psychologist, 17,* 300-319. https://doi.org/10.1027/1016-9040/a000102

Fergusson, D. M., & Horwood, L. J. (1997). Gender differences in educational achievement in a New Zealand birth cohort. *New Zealand Journal of Educational Studies, 32,* 83–96.

Flanagan, D., & Alfonso, V. (2017). *Essentials of WISC-V Assessment*. John Wiley & Sons.

Flanagan, D. P., & McDonough, E. M. (2018). *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (Fourth Edition). Guilford Press.

Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (pp. 1–7). Tilburg, The Netherlands: Tilburg University Press.

Gottfredson, L. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132. https://doi.org/10.1016/S0160-2896(97)90014-3

Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology/Psychologie canadienne, 50,* 183. https://doi.org/10.1037/a0016641

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29.0). Armonk, NY: IBM Corp

Kajonius, P. J. (2014). Honesty–Humility in contemporary students: Manipulations of self-image by inflated IQ estimations. *Psychological Reports*, *115*, 311–325. https://doi.org/10.2466/17.04.pr0.115c13z8

Kanazawa, S. (2013). Why is intelligence associated with stability of happiness? *British Journal of Psychology, 105,* 316-337. https://doi.org/10.1111/bjop.12039

Kelley, K., & Lai, K. (2012). MBESS: MBESS. R package version 3.3.2. Retrieved December 13, 2023, from http://CRAN. R-project.org/package=MBESS

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence, 36,* 153–160. https://doi.org/10.1016/j.intell.2007.03.005

Kort, W., Schittekatte, M., Compaan, E. L., Bosmans, M., Bleichrodt, N., Vermeir, G., Resing, W. & Verhaeghe, P. (2002). *Handleiding WISC-III^{NL}*. London, Amsterdam: The Psychological Corporation, NIP Dienstencentrum.

Kristjánsdóttir, D., & Zaiter, A. (2023). *Public domain intelligence tests: Psychometric properties of the Cog15 and ICAR16 cognitive ability scales* [Master thesis, Lund University]. Retrieved October 26, 2023, from https://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=9125503&fileOId=9125512

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of statistical software, 82,* 1-26. https://doi.org/10.18637/jss.v082.i13

Kwaliteitscentrum voor Diagnostiek (2023). *Beoordeling van de WISC-V-NL voor gebruik in Vlaanderen.* Retrieved August 25, 2023, from https://portaal.kwaliteitscentrumdiagnostiek.be/actueel/beoordeling-van-de-wisc-v-nl-voor-gebruik-in-vlaanderen/

Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: Insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews, 30,* 718-729. https://doi.org/10.1016/j.neubiorev.2006.06.001

Loe, B., Sun, L., Simonfy, F., & Doebler, P. (2018). Evaluating an Automated Number Series Item Generator Using Linear Logistic Test Models. *Journal of Intelligence, 6*, 20.

Low, L. K., & Cheng, H. J. (2006). Axon pruning: An essential step underlying the developmental plasticity of neuronal connections. *Philosophical Transactions of the Royal Society B-Biological Sciences, 361,* 1531-1544. https://doi.org/10.1098/rstb.2006.1883

Magez, W. (2019). *CoVaT-CHC Basisversie: Cognitieve VaardigheidsTest volgens CHC-model.* Centrum voor Psychodiagnostiek Thomas More. Retrieved August 25, 2023, from https://drive.google.com/file/d/1u4sXWgn5J0hhU_lk8xkWpAX9vjDZe60c/view

Magez, W., Tierens, M., Van Huynegem, J., Van Parijs, K., Decaluwé, V., & Bos, A. (2015). *CoVaT-CHC Basisversie.* Antwerpen: Psychodiagnostisch Centrum (PDC), Thomas More Hogeschool

Marcoulides, K. M., & Raykov, T. (2019). Evaluation of variance inflation factors in regression models using latent variable model ing methods. *Educational and Psychological Measurement, 79,* 874-882. https://doi.org/10.1177/0013164418817803

McDonald, R. P. (1999). *Test theory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum.

McLeod, J. W. H., & McCrimmon, A. W. (2021). Test Review: Raven's 2 Progressive Matrices, Clinical Edition (Raven's 2). *Journal of Psychoeducational Assessment, 39*, 388–392. https://doi.org/10.1177/0734282920958220

Merkle, E. C., Fan, J. & Zeileis, A. (2014). Testing for measurement invariance with respect to an ordinal variable. *Psychometrika, 79,* 569–584. https://doi.org/10.1007/s11336-013-9376-7

Ministry of Education and Training. (2012). *Omzendbrief SO/2012/01*. (SO/2012/01). Retrieved August 25, 2023, from http://data-onderwijs.vlaanderen.be/edulex/document.aspx?docid=14370

Ministry of Education and Training. (2022). Voltijds gewoon secundair onderwijs: algemeen overzicht (xlsx, 10 bladen) (150 kB) [Microsoft Excel spreadsheet]. Microsoft Corporation.

Ministry of Education and Training. (s.d.). *Naar het voltijds gewoon secundair onderwijs.* Retrieved December 1, 2023, from https://onderwijs.vlaanderen.be/nl/ouders/naar-school/naar-het-secundair-onderwijs/naar-het-voltijds-gewoon-secundair-onderwijs

Na, S., & Burns, T. (2016). Wechsler Intelligence Scale for Children-V: Test review. *Applied Neuropsychology: Child*, *5*(2), 156–160. https://doi.org/10.1080/21622965.2015.1015337

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution, 4,* 133-142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nelissen, W., Schittekatte, M., Fontaine, J., & Rigolle, F. (In preparation). *Validation of the International Cognitive Ability Resource (ICAR) in a Flemish sample*.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159. https://doi.org/10.1037/a0026699

Oakland, T., Douglas, S., & Kane, H. (2016). Top ten standardized tests used internationally with children and youth by school psychologists in 64 countries: A 24-year follow-up study. *Journal of Psychoeducational Assessment*, *34*(2), 166–176. https://doi.org/10.1177/0734282915595303

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41,* 673-690. https://doi.org/10.1007/s11135-006-9018-6

Pearson (2023). *Raven's 2 NL Scores invoeren.* Retrieved May 30, 2023, from https://qglobal.pearsonclinical.com/qg/static/Product/nl/index.htm#Ravens2-NL/Ravens2-NL_Enter_Scores.htm

R Core Team. (2024). *R: A language and environment for statistical computing* (version 4.3.3) [Software]. Vienna, Austria. Retrieved November 14, 2024, from http://www.R-project.org/

Raykov, T. (1998). Coefficient alpha and composite reliability with interrelated nonhomogeneous items. *Applied Psychological Measurement, 22,* 375–385. https://doi.org/10.1177/014662169802200407

Revelle, W., Condon, D., Wilt, J., French, J., Brown, A., & Elleman, L. (2017). Web- and phone-based data collection using planned missing designs. In N. Fielding, R. Lee, & G. Blank (Eds.), *The SAGE Handbook of Online Research Methods* (2nd ed., pp. 578–595). Sage Publishing.

Revelle, W., Dworak, E., & Condon, D. (2020). Cognitive ability in everyday life: The utility of open-source measures. *Current Directions in Psychological Science*, *29*(4), 358–363. https://doi.org/10.1177/0963721420922178

Rosseel, Y. (2012). Lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1-36. https://doi.org/10.18637/jss.v048.i02

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53,* 118-137. https://doi.org/10.1016/j.intell.2015.09.002

Sackett, P. R., Lievens, F., Van Iddekinge, C. H., & Kuncel, N. R. (2017). Individual differences and their measurement: A review of 100 years of research. *Journal of Applied Psychology, 102,* 254–273. https://doi.org/10.1037/apl0000151

Schmidt, F. L. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings.* ResearchGate. Available at https://doi.org/10.13140/RG.2.2.18843.26400

Schneider, W., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). The Guilford Press.

Seghers, M., Boone, S., & Van Avermaet, P. (2019). Classed patterns in the course and outcome of parent-teacher interactions regarding educational decision-making. *Educational Review, 73,* 417-435. https://doi.org/10.1080/00131911.2019.1662771

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., … Giedd, J. (2006). Intellectual ability and cortical development in children and adolescents. *Nature, 440,* 676-679. https://doi.org/10.1038/nature04513

Shaw, M., Rights, J. D., Sterba, S. S., & Flake, J. K. (2023). r2mlm: An R package calculating R-squared measures for multilevel models. *Behavior Research Methods, 55,* 1942-1964. https://doi.org/10.3758/s13428-022-01841-4

Stinissen, J., Smolders, M., & Coppens-Declerck, L. (1975). *Handleiding bij de Collectieve Verbale Intelligentietest voor derde en vierde leerjaar (CIT 3-4).* Brussel: C.S.B.O.

Tikhomirova, T., Malykh, A., & Malykh, S. (2020). Predicting academic achievement with cognitive abilities: Cross-sectional study across school education. *Behavioral Sciences, 10:*158. https://doi.org/10.3390/bs10100158

The International Cognitive Ability Resource Team. (2014). *International Cognitive Ability Resource.* Retrieved December 20, 2023, from https://icar-project.com

Vilia, P. N., Candeias, A. A., Neto, A. S., Franco, M. S., & Melo, M. (2017). Academic achievement in physics-chemistry: the predictive effect of attitudes and reasoning abilities. *Front. Psychol. 8:*1064. https://doi.org/10.3389/fpsyg.2017.01064

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition (WASI-II)* [Database record]. APA PsycTests. https://doi.org/10.1037/t15171-000

Weiss, L., Saklofske, D., Holdnack, J., & Prifitera, A. (Eds.). (2019). *WISC-V: Clinical Use and Interpretation.* Elsevier.

Wrulich, M., Brunner, M., Stadler, G., Schalke, D., Keller, U., & Martin, R. (2014). Forty years on: Childhood intelligence predicts health in middle adulthood. *Health Psychology, 33,* 292–296. https://doi.org/10.1037/a0030727

Young, S., & Keith, T. (2020). An Examination of the Convergent Validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, *38*(8), 1052–1059. https://doi.org/10.1177/0734282920943455

Young, S. R., Keith, T. Z., & Bond, M. A. (2019). Age and sex invariance of the International Cognitive Ability Resource (ICAR). *Intelligence, 77,* 101399. https://doi.org/10.1016/j.intell.2019.101399

Zinbarg, R., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and Mcdonald's ωH: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70,* 123–133. https://doi.org/10.1007/s11336-003-0974-7

Zorowitz, S., Chierchia, G., Blakemore, S. J., & Daw, N. D. (2023). An item response theory analysis of the matrix reasoning item bank (MaRs-IB). *Behavior Research Methods.* https://doi.org/10.3758/s13428-023-02067-8

Chapter 3.

Validating CalibrainVR-MR: Applying a VR-based mental rotation test to study orientation in secondary education

Submitted to Psychologica Belgica, April 2025.

Merel Dutry, Stijn Schelfhout, Eva Derous, Wouter Duyck, Klaas Bombeke, Jonas De Bruyne, Aleksandra Zheleva, Jelle Demanet, and Nicolas Dirix.

Abstract

Traditional 2D media-based mental rotation tests have been widely used to assess mental rotation ability. However, these tests come with inherent challenges, such as the ambiguity of 3D technical drawings and issues of occlusion. With the advent of immersive interfaces, virtual reality (VR) offers a promising alternative for more accurate assessments of mental rotation abilities, though its application in formal testing remains limited, particularly among younger populations. To address this void in literature, the present study evaluates the validity of CalibrainVR-MentalRotation (CalibrainVR-MR), a VR-based mental rotation test. The evaluation was conducted using a large sample of 1,223 students aged between eleven and fourteen who were tested for study orientation in secondary education. A subset of 131 students also provided longitudinal academic performance data, allowing for an in-depth analysis of the relationship between mental rotation abilities and academic performance in STEM versus non-STEM study environments. The results of the present study confirmed CalibrainVR-MR as a valid tool for research. Our findings also demonstrated significant correlations with STEM performance and showed no sex differences in the CalibrainVR-MR total score. We discuss the potential of CalibrainVR-MR as a supplementary tool in career education.

Introduction

Spatial ability is a multifaceted construct that involves the capacity to mentally represent and manipulate spatial information (Hegarty & Waller, 2005). This ability is essential for daily functioning, encompassing tasks such as navigating environments, arranging furniture, packing a suitcase, and accurately catching or throwing objects. Furthermore, spatial ability is important for proficiency and success in STEM (Science, Technology, Engineering, and Mathematics) disciplines (Sorby et al., 2018; Wai et al., 2009).

Despite a lack of general consensus regarding the exact constituents of spatial ability (Uttal et al., 2024), literature acknowledges mental rotation as a core component across all theories (Ariali & Zinn, 2021). Mental rotation ability refers to the capacity to rotate 2D or 3D figures and objects mentally (Linn & Petersen, 1985). This ability is commonly assessed using a mental rotation test. Two frequently used instruments are the Mental Rotation Test (MRT; Vandenberg & Kuse, 1978) and the Purdue Spatial Visualization Test: Rotations (PSVT:R; Guay 1979). While these 2D media-based assessments have been widely used, these tests come with inherent challenges, such as the ambiguity of 3D technical drawings and issues of occlusion (Bartlett & Camba, 2023). Recent technological advances offer promising solutions to enhance the assessment of mental rotation abilities. Specifically, recent research suggests that Virtual Reality (VR) may provide a more effective medium for this purpose (Bartlett et al., 2024).

Although VR and augmented reality (AR) technologies have been extensively employed in spatial skills training across various age groups (Piri & Cagiltay, 2024), their application in formal mental rotation ability testing remains limited (Bartlett et al., 2024). Existing studies validating VR-based mental rotation tests (Bartlett et al., 2024; Lochhead et al., 2022) primarily focus on adult populations, indicating a need for research validating these tests among younger populations, especially for use in educational contexts that require spatial abilities. This is quite rare, as test batteries to predict study choice, graduation or academic achievement are typically focused on, and often limited to, traditional cognitive skills like mathematics, linguistic skills or reasoning (e.g., SAT tests; College Board, 2019). Spatial reasoning is sometimes included as a component of a broader intelligence assessment, though usually it is omitted entirely. Therefore, such skills that are more important for technical (STEM) or vocational tracks remain underinvestigated (National Science Board, 2010).

The present study investigates the validity of a VR-based mental rotation test, CalibrainVR-MentalRotation (CalibrainVR-MR; Zheleva et al., 2023), in a sample of students aged between eleven and fourteen. By providing a more dynamic and realistic assessment environment, this test aims to overcome the limitations of existing 2D media-based tests. Moreover, the immersive nature of VR has the potential to make assessments more accessible by reducing test anxiety and fostering engagement through positive reactions (Mavridis & Tsiatsos, 2016; Mislevy et al., 2016). Given the strong evidence linking mental rotation abilities to success in STEM fields (Wai et al., 2009), we also investigate whether these associations can be established through VR-based mental rotation assessments.

Mental Rotation Ability

Mental rotation involves the ability to mentally rotate representations of two- and three-dimensional objects (Linn & Petersen, 1985). It is classified within the domain of spatial abilities, which constitutes a factor of human intelligence (Hegarty & Waller, 2005). Definitions of intelligence factors grounded in empirical evidence prove more robust (Buckley et al., 2018; Meehl, 2006); as such, in the present study we define spatial ability according to its empirically verified factor structure within the Cattell-Horn-Carroll (CHC) framework, which represents the most contemporary model of cognitive abilities (Schneider

& McGrew, 2012). Within the CHC framework, spatial ability is referred to as the Gv factor or visual processing[21] and is defined as: "the ability to generate, store, retrieve, and transform visual images and sensations" (McGrew, 2009, p.5).

Like other spatial and mathematical skills, individuals can enhance their mental rotation ability through practice and targeted training (Uttal et al., 2013). Research shows that mental rotation ability is important for success in STEM subjects (Sorby et al., 2018), and that improvements in mental rotation ability correlate with better performance in mathematics, computer science, natural sciences, and technology (e.g. Cheng & Mix, 2014; Sorby et al., 2013; Veurink & Sorby, 2019), as well as in vocational education (Lee, 2020). Moreover, incorporating spatial ability assessments into modern talent searches can help identify adolescents with untapped potential for STEM careers who might otherwise be overlooked (National Science Board, 2010; Wai et al., 2009). In the present study, we investigate whether the link between mental rotation ability and STEM performance also holds when measured using VR.

Assessment of Mental Rotation Ability

Mental rotation tests are widely used to assess mental rotation ability. However, traditional tests like the MRT (Vandenberg & Kuse, 1978) and PSVT:R (Guay, 1979) display 3D stimuli in a 2D format, which leads to visual issues that introduce construct-irrelevant variance and compromise the validity of these assessments (Bartlett & Camba, 2023). For instance, parts of the object may be occluded, and figures may be mistaken for 2D representations instead of 3D shapes (Bartlett & Camba, 2022; Caissie et al., 2009). Real-world depth cues such as shading, motion, and texture are crucial for interpreting 3D objects correctly (Pizlo & Steinman, 2008). However, these cues are missing in traditional tests that rely on black-and-white line drawings (Bartlett & Camba, 2023). Moreover, despite efforts to

[21]Although both spatial ability and visual processing describe the same construct, spatial ability is the more commonly used name.

make test stimuli more realistic, methods that rely on 2D media continue to encounter the challenge of occlusion (Bartlett et al., 2024).

By incorporating depth cues and enabling real-time viewpoint adjustments, VR-based tests can significantly reduce occlusion issues and enhance depth perception, leading to a more realistic and valid evaluation of mental rotation abilities (Bartlett et al., 2024). This approach can mitigate construct-irrelevant variance and can improve test performance, as demonstrated by research integrating 3D visualization technologies into spatial assessments (e.g., Sanandaji et al., 2017; Fisher et al., 2018). Although the available research provides promising evidence, the evaluation of mental rotation abilities using VR technologies remains relatively underexplored, with limited implementations over the past decade (Bartlett et al., 2024). However, modern extended reality (XR) technologies provide unprecedented opportunities to immerse users in lifelike environments, creating a sensation of actually being "there" (Ariali & Zinn, 2021) and are becoming increasingly accessible and affordable (e.g. the Meta Quest 2; Rana et al., 2023). Furthermore, the VR format enables more objective assessment scores compared to traditional pen-and-paper tests, as the format (1) can record metrics like reaction time (Weiner & Sanchez, 2020), (2) has the potential to reduce test anxiety (Mavridis & Tsiatsos, 2016) and (3) helps address sex disparity in mental rotation ability (Larson et al., 1999).

Indeed, a substantial body of research on the assessment of mental rotation ability has focused on sex differences. Traditionally, men have outperformed women on mental rotation tasks (Voyer et al., 1995). Differences are documented from early childhood through adulthood and increase with age (Geiser et al., 2008; Levine et al., 1999). However, interestingly these differences seem to disappear when real objects are used (McWilliams et al., 1997) or when virtual reality presentations are utilized (Larson et al., 1999; Parsons et al., 2004), suggesting that the sex gap is an unnecessary by-product of the artificial use of 2D representations to assess 3D reasoning. Other studies have similarly reported an absence of sex differences when using VR systems, indicating that immersive 3D presentations may provide a more equitable assessment method without compromising validity—and even enhancing it relative to real-word object manipulation (Neubauer et al. 2010). This finding aligns with evidence that sex differences in spatial abilities may be influenced by task presentation methods (Kozhevnikov et al., 2015; Seurinck et al., 2004). By providing egocentric perspectives—where the viewpoint aligns with the individual's own position—and minimizing allocentric challenges that require understanding spatial relationships from an external viewpoint, VR technology presents new opportunities for fairer evaluations of mental rotation abilities (Bartlett et al., 2024).

CalibrainVR-MentalRotation

An example of a VR-based mental rotation test is the CalibrainVR-MR, one of two core components of the CalibrainVR tool (Zheleva et al., 2023). CalibrainVR was initially designed as a calibration tool aimed at integrating biofeedback into VR experiences. Given the variability of physiological responses among individuals, proper calibration is essential for accurately interpreting biometric signals on a personalized level, especially for training and assessment purposes. CalibrainVR engages users in scientifically validated cognitive tasks while simultaneously capturing eye tracking and heart rate data to establish individual user profiles. The application focuses on two key cognitive processes vital for training and assessment: cognitive load and mental rotation ability. For the present study, we investigate the mental rotation task in an educational context.

CalibrainVR-MR is based on the pioneering work of Shepard and Metzler (1971), who developed the first mental rotation test involving the comparison of two simultaneously presented stimuli. In this pioneer test, the right stimulus was a rotated and/or mirrored version of the left stimulus, with varying angular disparities. Building on this approach, CalibrainVR-

MR immerses participants in a virtual factory environment where they are tasked with comparing pairs of stimuli (Zheleva et al., 2023).

Present Study

In the present study, we investigate the validity of CalibrainVR-MR among students within the Flemish Talentcenter framework (https://talentcenter.vlaanderen/). The Talentcenter helps students aged eleven to fourteen explore their interests, talents, and potential educational paths through various assessments and tasks, supporting their study orientation toward secondary education. Although students in Flanders do not formally select a study program during the first two years of secondary education, they must choose a school, a decision that can shape future opportunities, as study program offerings vary across schools. Additionally, students make preliminary track decisions, such as selecting elective courses, that prepare them for specific options in the third year. Talentcenter feedback supports both these early decisions and the program selection in the third year.

To evaluate the validity of CalibrainVR-MR, we conduct three key investigations. First, we assess the internal consistency of the test to ensure it reliably measures mental rotation. Second, we evaluate the test's construct validity by analyzing the pattern in test scores. Specifically, we investigate whether item difficulty aligns with the degree of rotation (Bartlett et al., 2024). Additionally, we examine its association with cognitive ability (McGrew, 2009) and its ability to replicate well-established relationships with actual STEM school performance (Ariali, 2020), thereby evaluating its relevance and potential impact in educational contexts. Finally, we examine potential sex differences[22] (Bartlett et al., 2024; Neubauer et al. 2010).

[22]In the present study, sex was used as a binary variable (male/female), based on participants' self-reported classifications. This categorization aligns with conventional definitions of sex, which are typically assigned at birth and associated with biological attributes such as chromosomal genotype, hormonal levels, and anatomy. While our study employed this binary approach, we acknowledge that sex is a complex spectrum that includes intersex individuals and individuals with differences in sex development.

Method

Ethical Statement

The research project[23] encompassing the Calibrain study obtained approval by the Ethical Commission of the Faculty of Psychology and Educational Sciences of Ghent University (reference number 2022/159). The participants and their parents gave their informed consent for participation, as well as for storage and use of the data by the researchers.

Participants and Procedure

The present study draws upon the secondary use of data gathered through the Talentcenter project, including records of students who visited the Talentcenter between May 2023 and December 2023 (N = 1,223). Of the participants, 49% identified as male, 33% as female, while 17% did not specify their sex. Additionally, 35% were in their final year of regular primary education (aged 11-12), 35% were in their first year of regular secondary education (aged 12-13) and 30% were in their second year of regular secondary education (aged 13-14).

Subsequent to the initial dataset compilation, additional data pertaining to academic performance and study choice during academic year 2023-2024 were collected. This supplementary data specifically pertained to students who had undergone the CalibrainVR-MR test during the preceding academic year (2022-2023) while being enrolled in the second year of secondary education. This subgroup is relevant, because these students were the only ones who had selected a specific study program at the time of the present research. Consequently, follow-up data were added for 131 students[24] of whom 62 chose for a STEM

[23] Entitled: "Pilot project of the Talentcenter for students in secondary education".

[24] In Flanders, from the third year of secondary education onwards, each study program belongs to a finality and domain of study (Onderwijskiezer, 2022). There are eight study (or interest) domains: Language and Culture, STEM, Art and Creation, Agriculture and Horticulture, Economy and Organization, Society and Well-being,

track, and 69 did not. Table 1 provides a detailed overview of the demographics and associated statistics of both the full sample and the longitudinal subsample, including data on the socio-economic status (SES) of participants in each group.

Table 1

Demographic Information Full Talentcenter Sample and Longitudinal Subsample

	Full Talentcenter sample	Longitudinal subsample
N	1,223	131
N female (%)	408 (33%)	49 (37%)
Grade		
Last year of primary education (%)	428 (35%)	0 (0%)
1st year SE (%)	429 (35%)	0 (0%)
2nd year SE (%)	366 (30%)	131 (100%)
Socio-economic status		
N home language is not Dutch (%)	112 (9%)	3 (2%)
*N r*eceiving a bursary (%)	352 (29%)	15 (12%)
N highest attained degree is primary education or lower Parent 1 (%)	80 (7%)	0 (0%)
N highest attained degree is primary education or lower Parent 2 (%)	118 (10%)	4 (3%)
N schools	33	3
M number of students per school *(SD)*	37.06 (45.20)	43.67 (50.62)
Number of students per school: range	8- 237	2 - 100

Sport, and Food and Catering; and three finalities. A finality indicates what students can do after secondary education: the "academic finality" prepares students for higher education, the "labor market finality" prepares for a job and the "double finality" prepares for both. Due to the distinctive curriculum and grading system associated with study programs belonging to the labor market finality, data pertaining to academic performance of students following these programs (*N* = 26) were omitted from the analyses. In the final longitudinal subsample, 16 participants were enrolled in a double finality program, and 115 in an academic finality program.

N in STEM-track (%)	62 (47%)
N girls in STEM (%)	22 (36%)
N receiving a bursary in STEM (%)	8 (13%)

Participants were required to complete the CalibrainVR-MR test as part of their Talentcenter experience. The test was administered individually, with two to three participants tested simultaneously. Before the test began, the Talentcenter coordinator explained the procedure, assisted participants in putting on the Meta Quest 2 headset, and guided them in holding the controllers, specifying that only the right hand was needed for responses.

Materials

CalibrainVR-MentalRotation

CalibrainVR-MR prompts participants to compare paired stimuli in a virtual environment and determine whether each pair contains identical stimuli, albeit rotated (Zheleva et al., 2023). The test involves comparing a 3D object, created with an in-application 3D printer, with a target image displayed on a screen mounted on the machine (Figure 1). This method integrates both 2D and 3D presentation formats. Participants are instructed to press a green button if they perceive the target and stimulus as identical or a red button if they perceive the target and stimulus as different. They are given a 7-item practice session before engaging in two blocks of 20 stimuli each. Block 1, categorized as the 'easy' block, involves stimuli rotated 10-30 degrees, whereas Block 2, classified as the 'difficult' block, involves stimuli rotated 150-170 degrees. Participants are free to move around, to observe the objects from multiple perspectives. The instructions emphasize the importance of being both quick and accurate in their responses. The total duration of the task is about 10 minutes (Zheleva et al., 2023).

Figure 1

First Person's Perspective of a Mental Rotation Item in CalibrainVR-MR

During the task, participants receive auditory and visual feedback to indicate their performance (Zheleva et al., 2023). For instance, the color of the factory walls changes to green or red depending on whether their answers are correct or incorrect, and a corresponding sound plays for each response. The accuracy of each response is recorded, with correct responses scored as one and incorrect responses as zero. To mitigate the impact of rapid guessing, correct responses that were given too quickly for thorough consideration were considered invalid: In our study this threshold was defined as correct responses occurring in less than 2.4 seconds based on the percentile threshold approach (Rios & Deng, 2021). The Calibrain total score was calculated as the proportion of correct responses.

STEM Performance

Mathematical Ability. Basic mathematical skills are assessed at the Talentcenter using a shortened version of the Cognitive Developmental Skills in Arithmetics (Cognitieve Deelhandelingen van het Rekenen [CDR]; Desoete & Roeyers, 2006). The CDR consists of 18 items and covers nine basic mathematical subskills that are required to calculate fluently (for a detailed description, we refer to Desoete & Roeyers, 2006). The test showed good internal consistency in our sample (α = .73; Evers et al., 2010).

Academic Grades (only for the longitudinal subsample). Academic grades include the Grade Point Average (GPA) and individual scores for mathematics and science courses. Individual course grades reflect a weighted average of scores on tests, assignments, and exams specific to each course, on a scale from 0 to 100. The GPA represents a weighted average of grades across all courses. All grades used in the present study reflect academic performance throughout the first trimester of academic year 2023-2024 (September to December), and were reported by the school principal in January 2024.

Due to variation in the number of science courses based on the student's study program, a composite variable was created to represent their average performance in science courses. This composite variable either reflects the score for a single science course, namely Natural Sciences, or is the average of scores in physics, biology, and chemistry.

Cognitive Ability

Cognitive ability is assessed at the Talentcenter using the Children's ICAR (Ch-ICAR; (Dutry et al., 2025; The International Cognitive Ability Resource Team, 2014). The Ch-ICAR consists of eight Verbal Reasoning items, six Matrix Reasoning items, seven Number Series items, and six Figural Analogies items. Verbal Reasoning items encompass multiple choice questions related to vocabulary, general knowledge, and mathematical problems. Matrix Reasoning items present eight geometric shapes arranged in a 3-by-3 grid, with one shape missing, requiring participants to select the correct ninth shape from the provided options. Number Series items display a short sequence of numbers, prompting participants to continue the sequence by entering the next number. Figural Analogies items feature three colored geometric shapes, where participants must select the missing fourth shape that fits the same way with the third shape as the second shape fits with the first one. An example of each item type is provided in Appendix 3A. Responses to the Ch-ICAR items are dichotomized: Correct answers are assigned a score of one, while incorrect answers receive a score of zero (Dutry et al., 2025). Subtest scale scores are derived by summing the binary values of each item within the subtest, while the Ch-ICAR total score is obtained by summing the subtest scale scores. In our sample, the full Ch-ICAR showed good internal consistency ($\alpha = .83$; Evers et al., 2010).

Control Variable

Socio-Economic Status. To determine the socio-economic status (SES) of the participants, we administered a questionnaire to the parents featuring three official indicators that are used by the Flemish Department of Education to allocate additional resources to schools with a high level of children low in SES (Ministry of Education and Training, 2012): parental financial capacity, educational attainment of both parents, and home language. Participant characteristics indicative for low SES are: receiving a scholarship, having one or two parents with a highest educational attainment of primary education or below, and speaking a language other than Dutch at home.

Analyses

Analyses were conducted in SPSS (version 29.0; IMB Corp., 2022) and R (version 4.3.3; R Core Team, 2024). All R analysis codes as well as the variance-covariance matrices are available at: https://osf.io/4evwt/?view_only=ff58b13d833f435aa6ab6b4f551708a4

To assess the internal consistency of CalibrainVR-MR, we calculated a Cronbach's alpha. To interpret the reliability coefficients, the COTAN guidelines for test quality evaluation are used (Evers et al., 2010). The COTAN rating system provides established criteria for evaluating reliability coefficients, tailored to the intended purpose of the test. COTAN identifies three primary purposes: tests for critical individual-level decisions, tests for less critical individual-level decisions, and tests for group-level research (Evers et al., 2010). For these purposes, reliability coefficients are considered sufficient if they fall between .80 and .90 for critical individual decisions, between .70 and .80 for less critical individual decisions, and between .60 and .70 for group-level research.

We evaluate the pattern of test scores by plotting a histogram of the CalibrainVR-MR total scores. To investigate the relationship between item difficulty and the degree of rotation, we computed the Spearman's rank correlation between the correct response rates and the degree of rotation. To explore the associations between CalibrainVR-MR performance, cognitive ability and STEM achievement, we first calculated Pearson correlations. Based on significant correlations, we then conducted linear regression analyses. Additionally, we compared CalibrainVR-MR performance between students in STEM and non-STEM study paths using an independent samples t-test. We also compared the strength of the Pearson correlations between CalibrainVR-MR and measures of academic achievement between both groups.

To investigate potential sex differences, we conducted an independent samples t-test. Additionally, we performed a Differential Item Functioning (DIF) analysis to detect potential sex differences at the item level, using the difR package (version 5.1; Magis et al., 2010). For the DIF analysis, we used the Mantel-Haenszel method with continuity correction and the CalibrainVR-MR score as matching variable for DIF detection.

Results

Internal Consistency

CalibrainVR-MR initially showed a Cronbach's alpha of .61. Using the "if item deleted" function in the Cronbach's alpha analysis, we evaluated the impact of each item on overall reliability, leading to the removal of six items (items 24, 29, 30, 31, 33, and 34). This refinement increased Cronbach's alpha to .70. This improvement not only reaches the threshold for good internal consistency required for research purposes, but also makes the measure suitable for individual-level decisions (Evers et al., 2010), thereby broadening its applicability and utility. Because of this substantial increase, the 34-item subset of

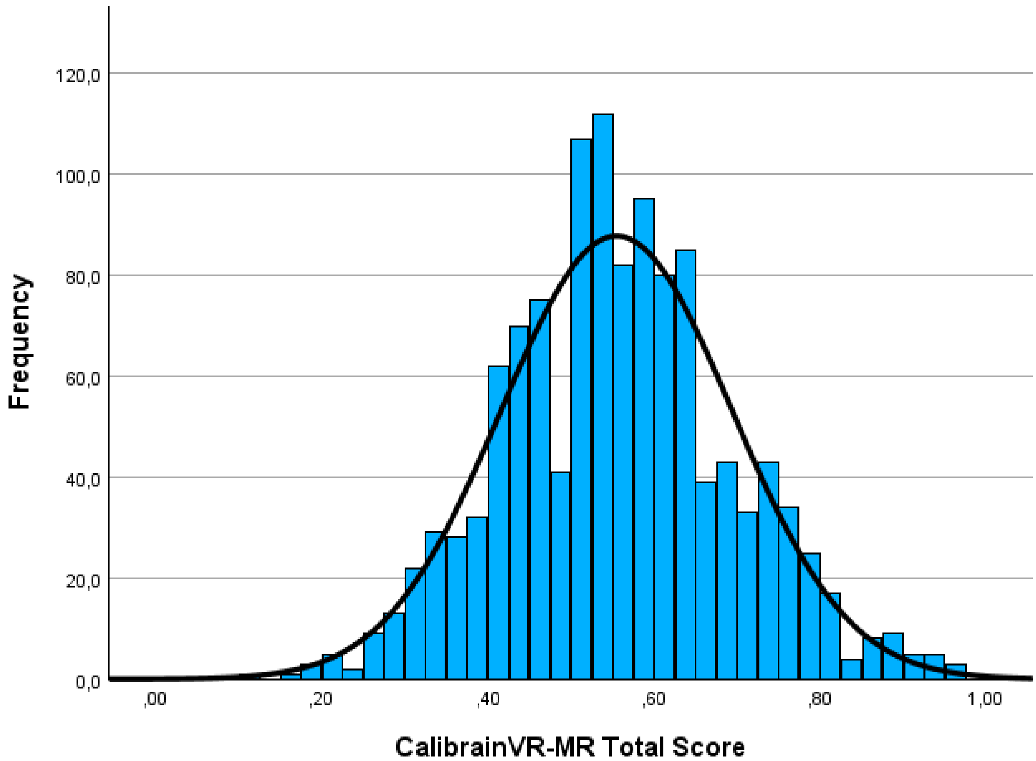
CalibrainVR-MR is used in further analyses. Appendix 3B gives an overview of the item parameters.

Construct Validity

The distribution of the CalibrainVR-MR total scores, representing the proportion of correct responses, is depicted in Figure 2. The distribution approximates a normal distribution, centered around a mean of $M = .55$ with a standard deviation of $SD = .14$.

Figure 2

Distribution of CalibrainVR-MR Total Scores



Note. The x-axis represents the proportion of correct responses

The Spearman's rank correlation analysis revealed a significant[25] negative correlation between the correct response rate and the degree of rotation: $r = -.43$, $p = .01$.

The CalibrainVR-MR total score demonstrated significant correlations with both cognitive ability and STEM-related performance measures. Specifically, the observed

[25]Due to the presence of ties in the data, the p-value was estimated using an approximate method.

correlation with cognitive ability, as measured by the Ch-ICAR, was $r = .33$, $p < .01$ (N = 1,223). After correcting for attenuation (Hedge et al., 2018; Muchinsky, 1996), this correlation increased to $r = .43$.

Similarly, the CalibrainVR-MR total score correlated with basic mathematical skills, as measured by the CDR, with an observed correlation of $r = .30$, $p < .01$ (N = 1,223). This correlation increased to $r = .42$, after correcting for attenuation (Hedge et al., 2018; Muchinsky, 1996).

In the longitudinal subsample (N = 131), all academic performance measures, including GPA, mathematics and science grades, showed positive and significant correlations with the CalibrainVR-MR total score (Table 2). To further investigate these relationships, we performed three linear regression analyses, with GPA, mathematics and science grades as the dependent variables. Each regression model included the CalibrainVR-MR total score as predictor, while controlling for sex, school and educational level parent 2, as this SES indicator was the only one to exhibit significant correlations with academic achievement (see Appendix 3C for the full correlation matrix). As shown in Table 3, all regression models were statistically significant, with the CalibrainVR-MR total score explaining 5%, 6% and 4% of the variance in GPA, mathematics and science grades respectively.

Table 2

Correlations Between CalibrainVR-MR and Academic Achievement

	1.	2.	3.	4.
1. Calibrain	1			
2. GPA	$.26^{**}$	1		
3. Math	$.26^{**}$	$.85^{**}$	1	
4. Sciences	$.24^{**}$	$.83^{**}$	$.71^{**}$	1

Note. $^{**}p < .01$. N = 131. *Calibrain* represents the CalibrainVR-MR total score.

Table 3

Results of Three Linear Regression Analyses

	GPA: $F(4,93) = 6.06$, $p < .001$, adjusted $R^2 = .17$				Mathematics: $F(4,93) = 4.05$, $p = .004$, adjusted $R^2 = .11$				Sciences: $F(4,93) = 3.06$, $p = .020$, adjusted $R^2 = .08$			
Predictor	B	*t*	*p*	Unique contribution	B	*t*	*p*	Unique contribution	B	*t*	*p*	Unique contribution
Intercept	62.01	19.07	<.001	-	46.34	7.76	<.001	-	61.12	12.86	<.001	-
CalibrainVR-MR	13.10	2.61	.01	5%	25.46	2.77	<.01	6%	16.01	2.19	.03	4%
Sex	3.31	2.47	.02	2%	1.15	.47	.64	-	2.22	1.13	.26	-
Education Parent 2	-.46	-.11	.91	-	-1.75	-.23	.82	-	-6.98	-1.15	.25	-
School	-12.22	-2.39	.02	4%	-18.28	-1.94	.05	3%	-7.06	-0.94	.35	-

Note. B represents the unstandardized coefficients. Since CalibrainVR-MR reflects the proportion of correct responses, B values for CalibrainVR-MR indicate the increase in the dependent variable corresponding to a full-scale increase from 0 to 1 in the CalibrainVR-MR total score.

To compare CalibrainVR-MR performance between students in STEM and non-STEM study programs, we conducted an independent samples t-test with STEM choice as the grouping variable and the CalibrainVR-MR total score as the dependent variable. The results indicated no significant difference in the CalibrainVR-MR total score between non-STEM students (M = .61, SD = .12) and STEM students (M = .64, SD = .14): $t(118.16) = -1.11$, p = .27, *Cohen's d* = -.20. Additionally, Pearson correlations between the CalibrainVR-MR total score and academic performance measures were calculated separately for students enrolled in STEM programs and those not in STEM. Results indicated stronger relations for students in STEM, with differences in correlation strength ranging from $\Delta r = .12$ to $\Delta r = .23$ (Table 4).

Table 4

Correlations Between CalibrainVR-MR and Academic Achievement in STEM and non-STEM Students

	Calibrain in non-STEM students N = 69	Calibrain in STEM students N = 62
GPA	.12	$.34^{**}$
Math	.13	$.36^{**}$
Sciences	.16	$.28^{*}$

Note. $^{*}p < .05$, $^{**}p < .01$. N = 131. *Calibrain* represents the CalibrainVR-MR total score.

Sex Differences

The independent samples t-test with sex as grouping variable and the CalibrainVR-MR total score as dependent variable showed no significant difference in CalibrainVR-MR total score between males (M = .56, SD = .14) and females (M = .55, SD = .13), $t(1009) = 1.22$, p = .22, *Cohen's d* = .08. Levene's test for equality of variances was not significant: $F(1,1009)$ = 1.87, p = .17), indicating that the assumption of equal variances is met.

The DIF analysis identified six items (items 7, 11, 13, 26, 37, and 39) displaying DIF

(Table 5). However, all items with DIF exhibited absolute effect sizes (log odds ratio) below 1.0, classified as negligible according to the ETS Delta scale (Holland & Thayer, 1988), indicating minimal to no practical impact (Magis et al., 2010). The DIF analysis results for all items are presented in Appendix 3D.

Table 5

Items Displaying DIF

Item n°	Mantel-Haenszel χ^2	*p*	log odds ratio effect size (DeltaMH)	DIF detected
7	4.99	.03	-.71 (A)	Yes
11	5.02	.03	-.57 (A)	Yes
13	5.07	.02	-.79 (A)	Yes
26	4.40	.04	.71 (A)	Yes
37	5.86	.02	.69 (A)	Yes
39	9.58	.00	-.92 (A)	Yes

Note. Effect size code A = negligible effect (Magis et al., 2010)

Discussion

The present study aimed to validate CalibrainVR-MR among students and assess its ability to replicate well-established relationships with STEM performance (Sorby et al., 2018; Wai et al., 2009). This validation is particularly important due to the current lack of validated VR applications designed for the formal assessment of mental rotation abilities, especially in younger populations (Bartlett et al., 2024; Lochhead et al., 2022; Weiner & Sanchez, 2020). Additionally, the available research suggests that VR may be a more effective medium for evaluating mental rotation abilities than traditional 2D media-based tests, offering a more dynamic and realistic assessment environment (Bartlett et al., 2024). Notably, sex differences commonly observed in mental rotation tasks tend to disappear when using VR systems (Neubauer et al., 2010), highlighting VR's potential to provide more equitable assessments. To assess the validity of CalibrainVR-MR, we evaluated its internal consistency, the construct validity and potential sex differences.

Regarding the internal consistency, the optimal Cronbach's alpha appeared after removing six items, reaching the threshold for good internal consistency required for research purposes and sufficient internal consistency for individual-level decisions (Evers et al., 2010). We thus utilized the 34-item subset in subsequent analyses and recommend that future research also adopt this subset when using CalibrainVR-MR.

Concerning score patterns, the score distribution closely approximated a normal distribution. This finding supports the test's quality, aligning with the expectation that cognitive ability factors typically follow the well-known bell curve (Flanagan & McDonough, 2018). Additionally, we observed a significant negative correlation between the degree of rotation and item difficulty, though the relationship was not perfect. This suggests that factors beyond the ability to mentally rotate objects are influencing the outcomes, like features of the shapes (Caissie et al., 2009). This finding is consistent with challenges typically encountered in mental rotation tests, regardless of the presentation format, although these challenges are less pronounced in VR settings (Bartlett et al., 2024).

Regarding the external validity, we found a medium correlation between CalibrainVR-MR performance and basic mathematical skills. This aligns with previous research showing that individuals with strong mental rotation skills often achieve high performance in mathematical tasks (Xie et al., 2020). The correlation we found with basic mathematical skills is slightly stronger than what has been observed in previous research (e.g. Casey et al., 1995; Chen, 2022; Geary et al., 2000; Rodán et al., 2016). Differences across studies likely arise from the varying types of items used to assess mathematical abilities, as some math tasks are more influenced by spatial representations, and the ease of translating quantitative relations into spatial terms varies (Geary et al., 2000; Rodán et al., 2016; Xie et al., 2020). Additionally, certain spatial tasks may align more closely with specific mathematical tasks due to similar underlying processes (Mix & Cheng, 2012).

We also observed a medium correlation between CalibrainVR-MR performance and cognitive ability. While factorial studies indicate a strong relationship between spatial ability and general cognitive ability (McGrew, 2009; Wechsler, 2014), it is important to recognize that mental rotation is just one component of the broader construct Gv. Research focusing specifically on mental rotation tasks and cognitive ability reports similar correlations (e.g., Chen, 2022; Santamaría et al., 2005).

The relationships between CalibrainVR-MR and academic performance measures—GPA, math and science scores—were weaker, with CalibrainVR-MR accounting for only a small proportion of the variance (5%, 6% and 4% respectively). Several factors may contribute to these findings. The sample for this analysis was rather small ($N = 131$) and was restricted in range (the majority of students were enrolled in higher education preparatory programs), which may have made it more difficult to detect relationships. Additionally, as standardized tests were not in place in Flemish secondary education at the time of writing, GPA and subject scores are based on individual teacher assessments, potentially affecting the results. Furthermore, variations in student motivation could further influence academic performance. Despite these limitations, the use of longitudinal data adds robustness to our findings, even though the associations are modest. Furthermore, the finding that CalibrainVR-MR is more strongly associated with academic performance measures in STEM study programs compared to non-STEM programs confirms its external validity and aligns with existing literature, which consistently shows the importance of spatial abilities within STEM fields (Sorby et al. 2013; Uttal & Cohen, 2012; Veurinck & Sorby, 2019). The absence of a significant difference in CalibrainVR-MR total score between the groups suggests that while students across disciplines may possess similar levels of mental rotation ability, its relevance and impact is more pronounced in STEM environments. This observation suggests that students in non-STEM programs with comparable mental rotation abilities may also have the potential to excel in STEM fields, presenting an opportunity to attract additional talent to these fields. This suggestion aligns with Wai et al. (2009), who indicated that incorporating spatial ability into talent searches could help identify adolescents with untapped potential for STEM, who are currently being overlooked. However, current test batteries designed to predict (STEM) study choice, graduation, or academic achievement predominantly emphasize traditional academic skills, such as mathematics, linguistic abilities, and reasoning (e.g., SAT tests; College Board, 2019). Furthermore, there is a notable paucity of valid VR-based spatial assessments for secondary students. The present study introduced a valid VR-based mental rotation test that can serve as a supplementary tool within existing test batteries, thereby enhancing the recognition of STEM talents and enabling students to make more informed educational choices.

Regarding sex differences, boys and girls exhibited equal variance in the CalibrainVR-MR total score, and no significant sex difference was observed. These results align with existing literature, which suggests that although men have historically outperformed women on traditional mental rotation tasks involving 2D representations (Voyer et al., 1995), this difference seems to disappear when tasks involve 3D objects (e.g., Larson et al., 1999; Neubauer et al., 2010; Parsons et al., 2004). This finding highlights a potential limitation of using 2D representations to assess 3D reasoning, as they may introduce an artificial sex gap by requiring an additional cognitive step not relevant to real-world 3D competency. Conversely, assessments using immersive 3D environments, such as VR, may provide a more accurate and equitable measure of mental rotation ability by aligning more closely with real-world object manipulation (Neubauer et al., 2010; Ramirez et al., 2020).

Notably, our study employed a combination of 2D and 3D presentation formats, allowing participants to engage with spatial stimuli in multiple dimensions. While this approach partially addresses the limitations of 2D-only tasks, as it incorporates real-world spatial configurations through 3D objects, it still requires an additional cognitive step for interpreting 2D representations that may not be entirely relevant to real-world 3D competency. Nonetheless, the inclusion of a fixed 2D reference may help reduce working memory load by providing a stable point of comparison, while the 3D components enhance participants' spatial understanding (Bartlett et al., 2024). This combination could help participants align and compare objects more effectively, redistributing cognitive load and enhancing the assessment of their mental rotation abilities. Nevertheless, future research should compare tests using a combined approach with those using exclusively 2D or 3D presentation formats to investigate this hypothesis further.

While our findings provide preliminary evidence that CalibrainVR-MR may serve as a sex-neutral tool, we remain cautious in drawing firm conclusions until further research can rule out other potential explanations for the observed sex-neutral effect. It is possible that sex differences simply did not exist in our sample, as the literature remains inconclusive regarding the developmental onset of sex differences in mental rotation abilities (Jansen et al., 2018; Rahe & Jansen, 2023; Rahe & Quaiser-Pohl, 2023; Titze et al., 2010). Furthermore, recent studies increasingly fail to replicate traditional sex differences, even in non-3D contexts, suggesting that the sex gap in this domain may be narrowing (Rodán et al., 2016; Toth & Campbell, 2019). This trend may be partially explained by the increasing engagement of women with video games and other technologies, which have been shown to enhance spatial abilities (Rodán et al., 2016; Terlecki et al., 2011). Historically, men had greater gaming experience, which may have contributed to superior performance on spatial tasks. However, with the advent of modern gaming technologies, women now engage with video games at comparable levels, using similar consoles, and seeking out entertaining gaming experiences (Terlecki et al., 2011). This shift may be contributing to the narrowing of the sex gap in mental rotation performance.

Practical Contributions

CalibrainVR-MR shows promise as a valuable, sex-neutral tool for career education, offering a means to help students identify their mental rotation abilities and learn about academic programs where these skills are particularly valued. By providing these insights, students can be empowered to make more informed decisions about their academic careers (Fonteyne et al., 2017). Within the Talentcenter framework, CalibrainVR-MR could play an important complementary role in guiding students toward fields that align with their strengths.

Additionally, the absence of sex differences in CalibrainVR-MR scores enhances its utility, particularly in addressing the persistent sex disparity in STEM enrollment, with fewer girls opting for STEM pathways (Veldman et al., 2021). The awareness that they possess comparable spatial abilities, could encourage more girls to consider and pursue careers in STEM disciplines (Lennon-Maslin et al., 2024). This potential to support greater sex equity in STEM further underscores the value of integrating CalibrainVR-MR into educational and career counseling efforts.

While the findings of the present study demonstrate that the CalibrainVR-MR tool is associated with STEM performance, caution is needed in interpreting these results. Predicting STEM proficiency requires multiple measures for a comprehensive assessment (Kuncel & Hezlett, 2010). While the present study highlights the importance of practical skills like mental rotation abilities for STEM success, CalibrainVR-MR represents only one metric. Ultimately, a balanced approach that includes multiple evaluative tools and considers non-academic factors is necessary to truly support the development of adolescents' STEM capabilities (Tai et al., 2006).

Limitations and Further Research

Although the analysis of the relationship with academic achievement was based on a longitudinal subsample of over 100 participants, which is notable for this type of research, it was restricted to students primarily enrolled in higher education preparatory programs, thereby restricting the range of cognitive abilities represented. Future research should aim to include more diverse samples across various educational contexts to determine whether our results can be replicated.

CalibrainVR-MR currently offers only two response options: green (yes, the two figures are the same), or red (no, they are not the same). This design choice was made intentionally to minimize cognitive load on participants. However, this binary choice inherently provides a 50% chance of answering correctly by guessing. Despite this, the distribution of test scores approximating a normal curve suggests that participants are not simply guessing; rather, they are selecting incorrect answers they believe to be correct. If guessing were prevalent, we would expect to see a more skewed distribution, characterized by a higher concentration of scores around 50% and very few scores below 50% accuracy. Therefore, our analyses likely reflect genuine participant responses. Nonetheless, it may be beneficial to consider adding an "I don't know" option or to explore an alternative test design where participants choose from three or more figures, selecting the one that matches the target figure.

Additionally, future research on CalibrainVR-MR should compare Calibrain performance with traditional paper-and-pencil tests, as well as with other spatial ability assessments in both 2D and 3D formats. This comparison will help explore the convergent validity of CalibrainVR-MR and provide insights into the advantages and limitations of combining 2D and 3D approaches.

Conclusion

The present study supports the utility of CalibrainVR-MR as a measurement of mental rotation abilities in students aged between eleven and fourteen. The 34-item subset demonstrated good internal consistency for research, and the absence of significant sex differences supports the potential of CalibrainVR-MR as a sex-neutral tool. The test's significant correlations with STEM performance, particularly in STEM programs, further underscore its relevance. Our findings suggest that CalibrainVR-MR holds promise not only for research but also for practical applications, such as educational and career counseling.

References

Ariali, S. (2020). Training of mental rotation ability in virtual spaces. *Journal of Technical Education, 8,* 46–63.

Ariali, S., & Zinn, B. (2021). Adaptive training of the mental rotation ability in an immersive virtual environment. *International Journal of Emerging Technologies in Learning, 16,* 20–39. https://doi.org/10.3991/ijet.v16i09.18971

Bartlett, K. A., & Camba, J. D. (2022). Isometric projection as a threat to validity in the PSVT: R. Minneapolis, MN. In 2022 ASEE Conference & Exposition Proceedings. Available at https://peer.asee.org/40396.pdf

Bartlett, K. A., & Camba, J. D. (2023). The role of a graphical interpretation factor in the assessment of spatial visualization: A critical analysis. *Spatial Cognition and Computation, 23,* 1–30. https://doi.org/10.1080/13875868.2021.2019260

Bartlett, K. A., Palacios-Ibáñez, A., & Camba, J. D. (2024). Design and Validation of a Virtual Reality Mental Rotation Test. *ACM Transactions on Applied Perception, 21,* 5. https://doi.org/10.1145/3626238

Buckley, J., Seery, N., & Canty, D. (2018). A heuristic framework of spatial ability: A review and synthesis of spatial factor literature to support its translation into STEM education. *Educational Psychology Review, 30,* 947–972. https://doi.org/10.1007/s10648-018-9432-z

Caissie, A. F., Vigneau, F., & Bors, D. A. (2009). What does the Mental Rotation Test measure? An analysis of item difficulty and item characteristics. *Open Psychology Journal, 2,* 94–102. https://doi.org/10.2174/1874350100902010094

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31,* 697–705. https://doi.org/10.1037/0012-1649.31.4.697

Chen, L. L. (2022). Spatial Ability and Mathematics Achievement in Deaf Children: The Mediating Role of Processing Speed and Intelligence. *Journal of Developmental and Physical Disabilities, 34,* 399-415. https://doi.org/10.1007/s10882-021-09805-w

Cheng, Y. L., & Mix, K. S. (2014). Spatial Training Improves Children's Mathematics Ability. *Journal of Cognition and Development*, *15*, 2–11. https://doi.org/ 10.1080/15248372.2012.725186

College Board. (2019). *Validity of the SAT for predicting first-year grades and retention to the second year.* College Board.

Desoete, A., & Roeyers, H. (2005). Cognitive skills in mathematical problem solving in Grade 3. *British Journal of Educational Psychology, 75,* 119-138. http://dx.doi.org/10.1348/000709904x22287

Dutry, M., Vereeck, A., Duyck, W., Derous, E., Schelfhout, S., Szmalec, A., … Dirix, N. (2025). Validation of the Children's International Cognitive Ability Resource (Ch-ICAR). *Behavior Research Methods, 57.* https://doi.org/ 10.3758/s13428-024-02591-1

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Review System for Evaluating Test Quality.* Boom Uitgevers Amsterdam. Retrieved from https://psynip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf

Fisher, M. L., Meredith, T., & Gray, M. (2018). Sex differences in mental rotation ability are a consequence of procedure and artificiality of stimuli. *Evolutionary Psychological Science, 4,* 124–133. https://doi.org/10.1007/s40806- 017- 0120- x

Flanagan, D. P., & McDonough, E. M. (2018). *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (Fourth Edition). Guilford Press.

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence, 36,* 556-563, https://doi.org/10.1016/j.intell.2007.12.003

Guay, R. B. (1976). Purdue Spatial Visualisation Test: Rotations

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning, *Journal of Experimental Child Psychology, 77,* 337-353. https://doi.org/10.1006/jecp.2000.2594

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods 50,* 1166–86. https://doi.org/ 10.3758/s13428-017-0935-1

Hegarty, M., & Waller, D. A. (2005). Individual Differences in Spatial Abilities. In P. Shah & A. Miyake (Eds.), The Cambridge Handbook of Visuospatial Thinking (121–170). Cambridge University Press.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Lawrence Erlbaum Associates.

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29.0). Armonk, NY: IBM Corp

Jansen, P., Ellinger, J., & Lehmann, J. (2018). Increased physical education at school improves the visual-spatial cognition during adolescence. *Educational Psychology, 38*, 964–976. https://doi.org/10.1080/01443410.2018.1457777

Kozhevnikov, M., Cheng, L. R., & Kozhevnikov, M. (2015). Effect of Environment Immersivity on Encoding Strategies of Spatial Tasks. *Procedia Manufacturing, 3,* 5059–5066, https://doi.org/10.1016/j.promfg.2015.07.519

Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19,* 339-345. https://doi.org/10.1177/0963721410389459

Larson, P., Rizzo, A. A., Buckwalter, J. G., & Van Roo. A. (1999). Gender issues in the use of virtual environments. *Cyberpsychology & Behavior, 2,* 113–123. https://doi.org/10.1089/cpb.1999.2.113. PMID: 19178247

Lee, I. J. (2020). Using augmented reality to train students to visualize three-dimensional drawings of mortise-tenon joints in furniture carpentry. *Interactive Learning Environments, 28,* 930-944. https://doi.org/ 10.1080/10494820.2019.1572629

Lennon-Maslin, M., Quaiser-Pohl, C., & Wickord, L. C. (2024). Beyond numbers: the role of mathematics self-concept and spatial anxiety in shaping mental rotation performance and STEM preferences in primary education. *Frontiers in Education, 9,* 1300598. https://doi.org/10.3389/feduc.2024.1300598

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35,* 940–949. https://doi.org/10.1037/0012-1649.35.4.940

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498. https://doi.org/10.2307/1130467

Lochhead, I., Hedley, N., Çöltekin, A., & Fisher, B. (2022). The Immersive Mental Rotations Test: Evaluating Spatial Ability in Virtual Reality. *Frontiers in Virtual Reality, 3,* 5. https://doi.org/10.3389/frvir.2022.820237

Magis, D., Beland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42,* 847–862. https://doi.org/ 10.3758/BRM.42.3.847

Mavridis, A., & Tsiatsos, T. (2016). Game-based assessment: Investigating the impact on test anxiety and exam performance. *Journal of Computer-Assisted Learning, 33,* 137–150. https://doi.org/10.1111/jcal.12170

McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1-10. https://doi.org/10.1016/j.intell.2008.08.004

McWilliams, W., Hamilton, C. J., & Muncer, S. J. (1997). On mental rotation in three dimensions. *Perceptual and Motor Skills, 85,* 297−298. https://doi.org/ 10.2466/pms.1997.85.1.297

Meehl, P. (2006). The power of quantitative thinking. In N. Waller, L. Yonce, W. Grove, D. Faust, & M. Lenzenweger (Eds.), A Paul Meehl reader: Essays on the practice of scientific psychology (pp. 433–444). Mahwah, New Jersey: Erlbaum.

Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M. I., von Davier, A., & John, M. (2016). Psychometrics and game-based assessment. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 23–48). New York, NY: Routledge.

Mix, K. S., & Cheng, Y. L. (2012). The relation between space and math: developmental and educational implications. *Advances in Child Development and Behavavior,* 42, 197–243. https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Ministry of Education and Training. (2012). *Omzendbrief SO/2012/01*. (SO/2012/01). Retrieved from http://data-onderwijs.vlaanderen.be/edulex/document.aspx?docid=14370

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*, 63–75. https://doi.org/10.1177/0013164496056001004

National Science Board. (2010). *Preparing the next generation of STEM innovators* (pp. 10-33). National Science Foundation.

Neubauer, A. C., Bergner, S., & Schatz, M. (2010). Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence, 38,* 529–539. https://doi.org/10.1016/j.intell.2010.06.001

Onderwijskiezer. (2022). *SO: 2e graad vanaf 2021-2022.* Consulted on 27 February 2024, of https://www.onderwijskiezer.be/v2/secundair/sec_2graad.php

Parsons, T. D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J. G., & Rizzo, A. A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia, 42,* 555–562. https://doi.org/10.1016/j.neuropsychologia.2003.08.014

Piri, Z., & Cagiltay, K. (2024). Can 3-Dimensional Visualization Enhance Mental Rotation (MR) Ability?: A Systematic Review. *International Journal of Human-Computer Interaction, 40,* 3683-3698. https://doi.org/ 10.1080/10447318.2023.2196161

Pizlo, Z., Li, Y., & Steinman, R. M. (2008). Binocular disparity only comes into play when everything else fails; A finding with broader implications than one might suppose. *Spatial Vision, 21*, 495–508. https://doi.org/10.1163/156856808786451453

R Core Team. (2024). *R: A language and environment for statistical computing* (version 4.3.3) [Software]. Vienna, Austria. Retrieved from http://www.R-project.org/

Rahe, M., & Jansen, P. (2023). Does mindfulness help to overcome stereotype threat in mental rotation in younger and older adolescents?. *Psychological Research,* 87, 624–635. https://doi.org/10.1007/s00426-022-01666-y

Rahe, M., & Quaiser-Pohl, C. (2023). Can (perceived) mental-rotation performance mediate gender differences in math anxiety in adolescents and young adults? *Mathematics Education Research Journal, 35,* 255-279. https://doi.org/10.1007/s13394-021-00387-6

Ramirez, D. E. A., Blinch, J., & Gonzalez, C. L. R. (2020). An evaluation of visuospatial skills using hands-on tasks. *Experimental Brain Research, 238,* 2269-2277. https://doi.org/10.1007/s00221-020-05894-9

Rana, M., Norbisrath, U., Vainikko, E., & Rossi, B. (2023). Transforming affordable virtual reality headsets into effective learning environments. *2023 International Conference on Intelligent Metaverse Technologies & Applications,* 168-175. https://doi.org/10.1109/iMETA59369.2023.10294604

Rios, J. A., & Deng, J. (2021). Does the choice of response time threshold procedure substantially affect inferences concerning the identification and exclusion of rapid guessing responses? A meta-analysis. *Large-scale Assessments in Education, 9.* https://doi.org/10.1186/s40536-021-00110-8

Rodán A., Contreras, M. J., Elosúa, M. R., & Gimeno, P. (2016). Experimental But Not Sex Differences of a Mental Rotation Training Program on Adolescents. *Frontiers in Psychology, 7,* 1050. https://doi.org/10.3389/fpsyg.2016.01050

Sanandaji, A., Grimm, C., & West, R. (2017). Inferring cross-sections of 3D objects: A 3D spatial ability test instrument for 3D volume segmentation. *In Proceedings of the ACM Symposium on Applied Perception,* 1–4. https://doi.org/10.1145/3119881.3119888

Santamaría, P., Arribas, D., Pereña, J., & Seisdedos, N. (2005). *EFAI, Evaluación Factorial de las Aptitudes Intelectuales.* Madrid: Departamento I+D TEA Ediciones.

Schneider, J., & McGrew, K. (2012). The Cattell–Horn–Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (3rd ed., pp. 99–144). New York: Guilford Press.

Seurinck, R., Vingerhoets, G., De Lange, F. P., & Achten E. (2004). Does egocentric mental rotation elicit sex differences?. *Neuroimage, 23,* 1440–1449. https://doi.org/10.1016/j.neuroimage.2004.08.010

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences, 26,* 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes.' *Learning and Individual Differences, 67,* 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Tai, R. H., Liu, C. Q., Maltese, A. V., & Fan, X. (2006). Planning early for careers in science. *Science, 312,* 1143-1144. https://doi.org/10.1126/science.1128690

Terlecki, M. S., Brown, J., Harner-Steciw, L., Irvin-Hannum, J., Marchetto-Ryan, N., Ruhl, L., & Wiggins, J. (2011). Sex differences and similarities in video game experience, preferences, and self-efficacy: implications for the gaming industry. *Current Psychology, 30,* 22–33. https://doi.org/10.1007/s12144-010-9095-5

The International Cognitive Ability Resource Team. (2014). *International Cognitive Ability Resource*. https://icar-project.com

Titze, C., Jansen, P., & Heil, M. (2010). Mental rotation performance and the effect of gender in 4th graders and adults. *European Journal of Developmental Psychology, 7*, 432–444. https://doi.org/10.1037/0022-3514.69.5.797

Toth, A. J., & Campbell, M. J. (2019). Investigating sex differences, cognitive effort, strategy, and performance on a computerised version of the mental rotations test via eye tracking. *Scientific Reports,* 9, 19430. https://doi.org/10.1038/s41598-019-56041-6

Vandenberg, S. G., & Kuse. A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Mot. Skills, 47,* 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 57, pp. 147-181). Oxford, UK: Academic Press.

Uttal, D. H., McKee, K., Simms, N., Hegarty, M., & Newcombe, N. S. (2024). How Can We Best Assess Spatial Skills? Practical and Conceptual Challenges. *Journal of Intelligence, 12,* 8. https://doi.org/10.3390/ jintelligence12010008

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139,* 352–402. https://doi.org/10.1037/a0028446

Veldman, J., Van Laar, C., Thoman, D. B., & Van Soom, C. (2021). "Where will I belong more?": The role of belonging comparisons between STEM fields in high school girls' STEM interest. *Social Psychology of Education, 24*, 1363–1387. https://doi.org/10.1007/s11218-021-09663-6

Veurink, N. L., & Sorby, S. A. (2019). Longitudinal study of the impact of requiring training for students with initially weak spatial skills. *European Journal of Engineering Education, 44,* 153–163. https://doi.org/10.1080/03043797.2017.1390547

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250–270. https://doi.org/10.1037/0033-2909.117.2.250. PMID: 7724690.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835. https://doi.org/10.1037/a0016127

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) Technical and Interpretive Manual.* Pearson.

Weiner, E. J., & Sanchez, D. R. (2020). Cognitive ability in virtual reality: Validity evidence for VR game-based assessments. *International Journal of Selection and Assessment, 3,* 215-235. https:// 10.1111/ijsa.12295

Xie, F., Zhang, L., Chen, X., & Xin, Z. Q. (2020). Is spatial ability related to mathematical ability: a meta-analysis. *Educational Psychology Review, 32,* 113-155. https://doi.org/10.1007/s10648-019-09496-y

Zheleva, A., De Bruyne, J., Durnez, W., Van De Walle, S., Verreyken, S., Demanet, J., & Bombeke, K. (2023). CaliBrainVR: Using psycho-physiological measures to calibrate virtual reality training. *2023 15th International Conference on Quality of Multimedia e-Experience, QOMEX,* 123-126. https://doi.org/10.1109/QOMEX58391.2023.10178522

Chapter 4.

Optimizing Career Guidance in Secondary Education: Exploring Person-Environment Interest Fit with the SIMON-jr Interest Inventory

Submitted to Journal of Career Assessment, April 2025.

Merel Dutry, Stijn Schelfhout, Eva Derous, Wouter Duyck, and Nicolas Dirix.

Abstract

Vocational interests and person-environment (PE) interest fit are well-established constructs that influence academic and career outcomes, and are widely applied in career education. However, despite the essential educational and career-related decisions that students face in early adolescence, PE interest fit has been largely overlooked at this developmental stage, with most research focusing on older adolescents or adults. The present study addresses this gap by exploring the role of PE interest fit in early adolescence. We introduce the SIMON-jr questionnaire, a freely accessible tool designed to assess vocational interests in adolescents aged eleven to fourteen. Using the SIMON-jr, we examined how students' vocational interests in the first year of secondary education (ages 12-13) fit with their study program choices by the third year (ages 14-15), and how this fit relates to academic achievement. Additionally, we explored potential differences in PE interest fit between boys and girls. Our results confirmed the reliability and validity of SIMON-jr, establishing the tool as a robust and equitable resource for guiding students in their educational choices. Furthermore, our findings indicated that secondary students tend to select education programs that fit their interests, with girls showing a stronger interest fit than boys. At the same time, interest fit did not relate to academic achievement at this young age, for either gender. We address several reasons for this lack of association, and highlight the potential of career education tools that assess PE interest fit to support early adolescents in making informed academic and career decisions.

Introduction

The process of choosing a study path in secondary education is often highly individualized, with decisions expected to align with students' interests and skills (Nauta, 2010; Slot et al., 2021). However, in reality, students often make choices with limited information. Many twelve-year-olds (and their parents) lack a clear understanding of what a particular field of study entails, or underestimate the long-term consequences of their choices (Goosen et al., 2017). Moreover, study decisions are typically driven by presumed cognitive levels and academic demands, while actual domain interest often plays only a minor role. This tendency is concerning, as research indicates that a good fit between personal interests and the educational environment significantly benefits study completion, career satisfaction, and overall success (de Vries et al., 2024; Nye et al., 2012). Recognizing these challenges, the present study adopts an interest-oriented approach to explore its potential advantages in supporting student's educational choices in secondary education.

Vocational interest assessment and the evaluation of person-environment (PE) interest fit—the congruence between an individual's vocational interests and the interests sparked in the environment—have proven to be valuable tools in career education, particularly in studies involving adults and late adolescents (Fonteyne et al., 2017; Hanna et al., 2021; Schelfhout et al., 2021a). PE interest fit provides a structured framework that supports individuals in exploring options and engaging in reflective analysis of their personal preferences (Schelfhout et al., 2021b), two critical components of effective career education (Hughes et al., 2016; Kuijpers & Scheerens, 2006). However, despite the essential educational and career-related decisions that students face in early adolescence (Seghers et al., 2019), PE interest fit has been largely overlooked at this developmental stage (Su, 2020). Addressing this gap, the present study introduces the SIMON-jr (SIMON-junior), a newly developed, publicly accessible questionnaire, available in Dutch, designed to measure vocational interests in early adolescents (aged 11 to 14). Using SIMON-jr, we analyze the degree of PE interest fit between students' vocational interests in the first year of secondary education (ages 12-13) and their study choices by the third year (ages 14-15), as well as how this fit relates to academic achievement. By focusing on this age group, our findings aim to provide study counselors with valuable insights to improve decision-making guidance and support the development of career guidance tools tailored to the needs of early adolescents.

Vocational Interests

Research on PE interest fit often relies on Holland's theory of vocational interests (1997), which is still one of the most influential frameworks in vocational literature. The theory's central premise posits that individuals are more likely to thrive in educational and occupational settings that fit their interests. Holland identified six primary dimensions—Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC)—to classify both individuals' interests and their environments, allowing for consistent, commensurate measurement across both. Holland's theory has been empirically verified and to this day still inspires the development of numerous interest inventories, linking interests with academic subjects and career paths across various contexts (Fonteyne et al., 2017; Lasselle et al., 2021; Lavrijsen et al., 2021; Nauta, 2010).

The RIASEC interests are typically arranged in a hexagonal pattern, reflecting a circumplex structure (Tracey & Rounds, 1996). According to Holland's theory (1997), adjacent interest types on this hexagon exhibit stronger correlations (e.g. R-I, I-A, A-S) than those that are further apart (e.g. R-S, I-E, A-C; Nauta, 2010). This structure implies that as individuals score higher on one vocational dimension, they are likely to score lower on its opposing dimension (Nagy et al., 2010; Nauta, 2010).

However, recent literature in career counseling and vocational interests increasingly recognizes the complexity of interest structures, suggesting that individuals, especially in early adolescence, can hold multiple vocational interests simultaneously (Slot et al., 2021). Tay et al. (2011) were among the first to challenge this bipolarity principle of the RIASEC model, showing that individuals can have interests spanning opposing dimensions. This insight has spurred further research into how these diverse vocational interests combine to create distinct profiles (e.g., McLarnon et al., 2015; Perera & McIlveen, 2018; Sung et al., 2017). Consequently, it is recommended to incorporate all six RIASEC dimensions to evaluate PE interest fit rather than restricting analysis to the traditional three-letter Holland code (Slot et al., 2021; Xu & Li, 2020), in which only the dominant letters are retained for interest guidance and counseling.

PE Interest Fit

Commensurate RIASEC measurement of both persons (P) and environments (E) allows us to assess the PE fit between individual interests (e.g. of a secondary education student) and environmental characteristics (e.g., a study program) (Schelfhout et al., 2021a). To achieve commensurate measures, literature describes methods to obtain individual and environmental profiles. An individual's RIASEC profile is typically obtained by completing a RIASEC interest questionnaire, which provides scores across each of the six dimensions. To determine the profile of an environment, there are two commonly used methods (de Vries et al., 2024; Schelfhout et al., 2022). The *a priori* method assigns a RIASEC profile to a study program or occupation based on existing classifications such as O*NET or Holland Codes (Rounds et al., 1999; Rounds et al., 2021). The *sample-based* method generates a profile by calculating the average RIASEC scores of individuals within that environment. Here, individuals serve as representatives or *incumbents* of their chosen environment. Early research by Assouline and Meir (1987) already indicated that sample-based interest fit tends to show a stronger correlation with satisfaction than the a priori approach, supporting Holland's idea that people themselves shape their environments (Chartrand & Walsh, 1999).

Literature operationalizes PE interest fit using a number of different methods. Xu and Li (2020) categorized these different methods into three broad approaches: top-letter(s)-oriented, profile-based empirical, and profile-based conceptual. Top-letter(s)-oriented approaches assess fit by examining whether the highest-scoring RIASEC dimension(s) of an individual align with the primary characteristic(s) of the environment, focusing exclusively on dominant interest fit. In contrast, profile-based approaches assess fit across the entire RIASEC profile. The empirical profile-based approach uses polynomial regression to estimate fit, capturing nuanced information through linear, curvilinear, and interaction effects (Schelfhout et al., 2022; Shanock et al., 2010; Wiegand et al., 2021). Finally, the conceptual profile-based approach evaluates fit by relying on environmental profiles and employs methods like profile correlation (Xu & Li, 2020). Profile correlation correlates the relative ranking of RIASEC interests across person and environment profiles, producing a correlation coefficient between -1 (complete misfit) and 1 (perfect fit).

When choosing a method to assess interest fit, both predictive capability and interpretability are crucial (Granillo-Velasquez et al., 2024). Furthermore, it is essential to ensure that the method aligns conceptually and empirically with its intended use. In career education settings, where the goal is often to recommend educational or occupational careers to explore aligned with individual interests, research indicates that profile correlation may be the most useful operationalization due to its predictive strength and interpretability (Granillo-Velasquez et al., 2024; Xu & Li, 2020). Therefore, the present study adopts profile correlation as operationalization of PE interest fit.

PE Fit in Educational Contexts

Holland's (1997) theory suggests that people are attracted to environments that fit their interests and that individuals are more likely to thrive academically or professionally when working or studying in environments that match their interests (Hoff et al., 2020; Lent et al., 1994; Tracey et al., 2012). Recent meta-analyses show significant associations between PE fit and outcomes in both occupational (Hoff et al., 2020; Nye et al., 2017; Van Iddekinge et al., 2011) and educational contexts, including academic achievement, persistence, and satisfaction (de Vries et al., 2024; Nye et al., 2012). The strength of these associations varies depending on the conceptualization and measurement of vocational interests, with effect sizes of up to $r = .36$ (Nye, Butt et al., 2018; Nye, Prasad et al., 2018; Schelfhout et al., 2022).

Despite the established importance of vocational interests in shaping individuals' academic and professional development (Wille et al., 2020; Hidi, 2006; Rounds & Su, 2014) previous research has primarily focused on adult and higher education populations (de Vries et al., 2024; Su, 2020). Only recently however, researchers are starting to investigate the role of vocational interests in shaping academic outcomes within secondary school settings. For instance, Slot et al. (2021) found that vocational interests influence educational choices among lower secondary students (ages 14–15) in ways similar to college students' major selection (e.g., Pässler & Hell, 2012), particularly for students with dominant interests in specific dimensions. For example, students with an enterprising–conventional dominant profile predominantly selected the economics–society track, while those with a realistic–investigative dominant profile gravitated toward the nature–technology tracks. In contrast, students with non-differentiated profiles, including those with low, average, or broad interests across all dimensions, exhibited greater variability in their educational choices. Similarly, Lavrijsen et al. (2021) demonstrated that both vocational interests and cognitive abilities independently predict subject preferences in middle school students (ages 12-13), underscoring their significance in understanding students' educational and career paths. Lavrijsen et al. (2021) also highlighted how vocational interests can explain more nuanced distinctions between subject preferences. For example, students who scored high on the realistic interest dimension were more likely to prefer STEM subjects that integrate mathematics and science in applied contexts, while those with stronger investigative interests tended to favor traditional mathematics and science subjects. Related to these findings, Hyland et al. (2022) found that secondary students (ages 14-18) with interest–ability profiles closely aligned with specific knowledge areas consistently outperformed those with less alignment in terms of knowledge acquisition, even when the latter possessed higher overall ability levels. This finding underscores the vital role of vocational interests in determining not only the types of information students pursue but also the persistence and intensity with which they engage in that pursuit (Su, 2020).

Although the studies of Hyland et al. (2022), Lavrijsen et al. (2021) and Slot et al., (2021) highlighted the importance of assessing vocational interests in guiding secondary students, none have specifically examined PE interest fit between students and their secondary education programs, nor its relations with academic achievement. Addressing this gap is crucial, as interest theory suggests that the fit between a student's vocational interests and their educational environment is a stronger predictor of performance than interests alone (Nye et al., 2012).

Sex Differences in Vocational Interests and Interest Fit

Sex differences in vocational interests are well-documented and emerge early in development. Among adolescents and adults, women typically show stronger Social, Artistic, and Conventional interests, while men often exhibit stronger Realistic and Investigative interests (Su et al., 2009). These trends are already evident in late childhood and early adolescence, with girls reporting higher Social, Artistic, and Conventional interests and boys favoring Realistic interests (Maurice & Bäumer, 2015; Tracey, 2002). Findings regarding Enterprising and Investigative interests are more mixed at this stage (Pässler & Hell, 2020; Xu & Tracey, 2016). Interestingly, the magnitude of these differences seems to evolve with age. For example, the sex difference in Realistic and Social interests is smaller during late childhood and early adolescence compared to older samples, whereas the gap for Artistic interests is larger (Pässler & Hell, 2020; Tracey, 2002; Xu & Tracey, 2016).

In addition to specific vocational interests, sex differences are also observed in the degree of differentiation within interest profiles. Girls tend to display more differentiated profiles, showing clear preferences for specific domains over others (Fouad & Mohler, 2004; Slot et al., 2021; Sung et al., 2017). In contrast, boys more often exhibit nondifferentiated profiles, characterized by broad, neutral, or low-level interests that do not strongly favor any particular domain. Differentiated profiles are linked to a more precise alignment between interests and educational or career choices, whereas nondifferentiated profiles often lead to more divergent paths (Schelfhout et al., 2021a; Slot et al., 2021).

Present Study

The present study aims to enhance career education by exploring the role of PE interest fit in secondary education, using the SIMON-jr questionnaire as assessment tool. SIMON-jr is a newly developed, publicly accessible[26] questionnaire, available in Dutch, designed to measure vocational interests in early adolescents (aged 11 to 14). This instrument builds on two well-established foundations: the SIMON-I (Fonteyne et al., 2017), a validated interest inventory extensively used in Flemish higher education to guide tens of thousands of students in exploring academic and vocational pathways, and the ICA-R (Tracey & Ward, 1998), a U.S.-developed tool designed to assess RIASEC interests in children. By tailoring the SIMON-I framework to younger students, SIMON-jr extends a reliable career counseling tool to support students in secondary education. Additionally, drawing on the ICA-R, SIMON-jr ensures age-appropriate and theoretically grounded insights into early adolescents' vocational preferences.

The present study first examines the psychometric properties of SIMON-jr, focusing

---

[26]The items and scoring of the SIMON-jr are provided in Appendix 4A.

on the internal consistency of the six interest dimensions, the alignment with the RIASEC structure, and sex differences. While replicating the established sex differences with SIMON-jr would support the construct validity of the instrument, it is important to ensure that these differences reflect empirical disparities between boys and girls rather than sex biases in the newly developed interest scales, and also regardless of the biological or environmental origin of such observed differences (which is not the focus of this study). This psychometric evaluation aims to confirm the reliability and validity of SIMON-jr in measuring vocational interests among early adolescents, ensuring the instrument serves as a robust and equitable resource for guiding students in their educational choices.

Following the psychometric evaluation, we first analyze PE interest fit by examining how students' vocational interests in the first year of secondary education (ages 12-13) align with their study program choices three years later (ages 14-15). Second, we assess whether this PE interest fit predicts academic achievement. Considering the literature on PE interest fit, which emphasizes that fit between individual interests and environmental characteristics benefits both personal and academic outcomes (de Vries et al., 2024; Holland, 1997; Nye et al., 2012), we hypothesize that (Hypothesis 1) students' vocational interests will align with the vocational characteristics of their study program and (Hypothesis 2) PE interest fit will be positively associated with academic achievement, regardless of the specific study program. Additionally, grounded in research suggesting that girls tend to display more differentiated interest profiles than boys, which are associated with stronger fit between their interests and study programs (Schelfhout et al., 2021a; Slot et al., 2021), we hypothesize that (Hypothesis 3) girls will achieve stronger PE interest fit compared to boys.

Method

Context

As our study examines PE interest fit within the context of Flanders, we begin by providing a brief overview of the Flemish secondary education system. In Flanders, compulsory secondary education is structured into three two-year cycles with increasing levels of specialization. The first cycle comprises two primary streams: the A-stream and the B-stream. The B-stream is specifically designed for students who have not met the minimum requirements of primary education (Flemish Government, n.d.). In contrast, the majority of students—86% in 2023-2024—are enrolled in the A-stream (Department of Education and Training, 2023). Despite offering a standardized curriculum, the A-stream introduces early tracking through optional courses, with schools offering options that align with the study programs they provide from the second cycle onward (Seghers et al., 2021). In the second cycle, students select a study program from one of eight domains (e.g. Language and Culture, STEM, or Sport), along with one of three "finalities": academic (preparation for higher education), vocational (preparation for workforce entry), or dual (preparation for both). Additionally, twelve cross-domain programs are available within the academic finality. Individual study programs are classified within this domain-by-finality matrix based on content similarities. The present study followed students from the first (ages 12-13) to the third year (ages 14-15) of secondary education.

Participants

This study draws on data from the Talentcenter Study[27], a longitudinal project following students in Flanders through the first three years of secondary education. The sample consists of *N* = 868 students who contributed data to at least one wave of data

[27]Note that "Talentcenter" is a proper name.

collection used in this study. 10% of these participants reported a home language other than Dutch (10% unspecified). Additionally, 28% of the participants had parents who received a scholarship (financial status unspecified for 19%). In terms of parental education, 9% had one parent with a maximum education level of primary school, and 5% had both parents with only primary education or lower (unspecified for 11% of the participants).

During the first data collection (grade seven, September-November 2021), 833 students completed the vocational interest questionnaire (M_{age} = 12.02 years, SD_{age} = .55, 46% female). These students were enrolled in 23 different regular secondary schools across Flanders, with 730 students in the A-stream and 103 in the B-stream. During the final data collection (grade nine, January-February 2024), we obtained data on the chosen study programs of 606 students (M_{age} = 14.23 years, SD_{age} = .50, 45% female), spread across 36 different schools, with 582 also providing information on their academic achievement.

The research project[28] encompassing the present study was approved by the Ethical Commission of the Faculty of Psychology and Educational Sciences of Ghent University (reference number 2021/59). Both the parents of the participants and the participants themselves gave their informed consent for participation, as well as for storage and use of the data by the researchers.

Materials

Vocational Interests

The SIMON-jr inventory was used to assess students' vocational interests. This inventory was specifically designed for the Flemish educational context, targeting early adolescents (aged 11 to 14). SIMON-jr is grounded in Holland's (1997) RIASEC interest model and comprises 36 items, with 6 items representing each interest dimension. Each item describes either an activity (26 items) or an occupation (10 items) associated with a particular

[28] Entitled: "Study orientation in secondary education".

interest dimension. Participants respond with “yes” or “no” to indicate if they would enjoy the specified activity or occupation. For each interest dimension, the percentage of “yes” responses is calculated, and these percentages create the participant's RIASEC profile. Details on internal consistency and structure are provided in the results section.

Study Environment Interest Profiles

Using a sample-based approach, we established an interest profile for each cell in the Flemish matrix of study programs—classified by domain and finality based on content similarities—as well as separate profiles for each of the cross-domain programs (see also the ‘Context’ section). Creating interest profiles for grouped study programs, rather than for individual programs is appropriate, as most programs within a cell displayed similar interest profiles. Detailed research methods for defining these interest profiles are provided in Appendix 4B.

Academic Achievement

Academic achievement was measured using the Grade Point Average (GPA) from the first trimester (September to December), reflecting a weighted average of assignments, tests and exams across all subjects. GPA data were reported by the school principal. Since standardized testing is not implemented for third-year secondary students in Flanders, GPA scores are derived from individual teacher assessments. This system can introduce variability, as grading practices may differ between teachers and subjects.

Procedure

Participants filled out the interest questionnaire as part of the longitudinal Talentcenter Study. The survey, programmed in Qualtrics (https://www.qualtrics.com/), was conducted in the participants’ classroom during regular school hours. Participants completed the questionnaire on an internet-enabled device (laptop, computer or tablet), and the class teacher and a supervisor of Ghent University were present. After a brief explanation, the participants were able to work independently on the test. No time limits were imposed.

For follow-up data in the third year of secondary education, school principals provided information on students' chosen study programs and academic performance. If students had changed schools, they were contacted via mail to supply details about their chosen study program and academic achievement. Nevertheless, most data were obtained through school principals.

Analyses

Analyses were conducted in SPSS (version 29.0; IMB Corp., 2022) and R (version 4.3.3; R Core Team, 2024). All R analysis codes as well as the variance-covariance matrix are available at: https://osf.io/esf8y/?view_only=aae3a86696c543c380b01dd3b375201d.

To assess the validity of SIMON-jr, we first examined the internal consistency of the RIASEC dimensions using Cronbach's alpha. The reliability coefficients are interpreted following the COTAN guidelines for evaluating test quality (Evers et al., 2010). These guidelines outline specific reliability criteria tailored to the intended use of a test, categorized into three main purposes: high stakes decisions, less critical individual-level decisions, and group-level research. According to COTAN, reliability coefficients are deemed sufficient if they range between .80 and .90 for high stakes testing, .70 to .80 for less critical individual decisions, and .60 to .70 for group-level research (Evers et al., 2010).

Second, the RIASEC structure of the SIMON-jr questionnaire was evaluated through confirmatory factor analysis (CFA) in R using the lavaan package (version 0.6.17; Rosseel, 2012). A six-factor model was tested using Diagonally Weighted Least Squares (DWLS) estimation to account for the binary nature of the data (Xia & Yang, 2018). To examine the circular structure, we calculated Pearson correlations and conducted a Randomization Test of hypothesized Order Relations (RTOR), using the RANDALL package (Fonteyne et al., 2017; Hubert & Arabie, 1987; Tracey & Rounds, 1993).

Finally, to examine sex differences in vocational interests, we conducted Mann-Whitney U tests because the vocational interest data did not meet normality assumptions. A Bonferroni correction was applied to control for multiple testing, setting the adjusted significance threshold at $p < .0083$ (= .05/6). Furthermore, to ensure that the observed sex differences represent empirical disparities rather than item bias, we conducted a Differential Item Functioning (DIF) analysis using the difR package (version 5.1; Magis et al., 2010). For the DIF analysis, we used Mantel-Haenszel method with continuity correction and the interest score as matching variable for DIF detection.

The interest fit between person and environment was operationalized using profile correlation, which compares students' individual interest profiles to the interest profile of their chosen study program. These program profiles were generated using a sample-based approach, further detailed in Appendix 4B. To explore the association between PE interest fit and academic achievement, Pearson correlations were calculated. These analyses were also conducted separately for boys and girls to identify potential sex-specific patterns, with independent samples t-tests performed to assess whether differences in PE interest fit between boys and girls were statistically significant. To account for between-school variability in GPA grading practices, we also performed the same analyses on within-school standardized GPA scores.

Results

SIMON-jr

Internal Consistency

All RIASEC dimensions showed sufficient to good internal consistency for group-level research (Evers et al., 2010). The Cronbach's alphas were, corresponding to the order of the RIASEC-acronym: .80, .83, .82, .79, .70 and .62.

Structure

The distribution of interests scores across each of the RIASEC dimensions is illustrated in Figure 1. Means and standard deviations for each dimension, with separate values for boys and girls, are presented in Table 1.

Figure 1

Distribution of Interests Scores across the RIASEC Dimensions

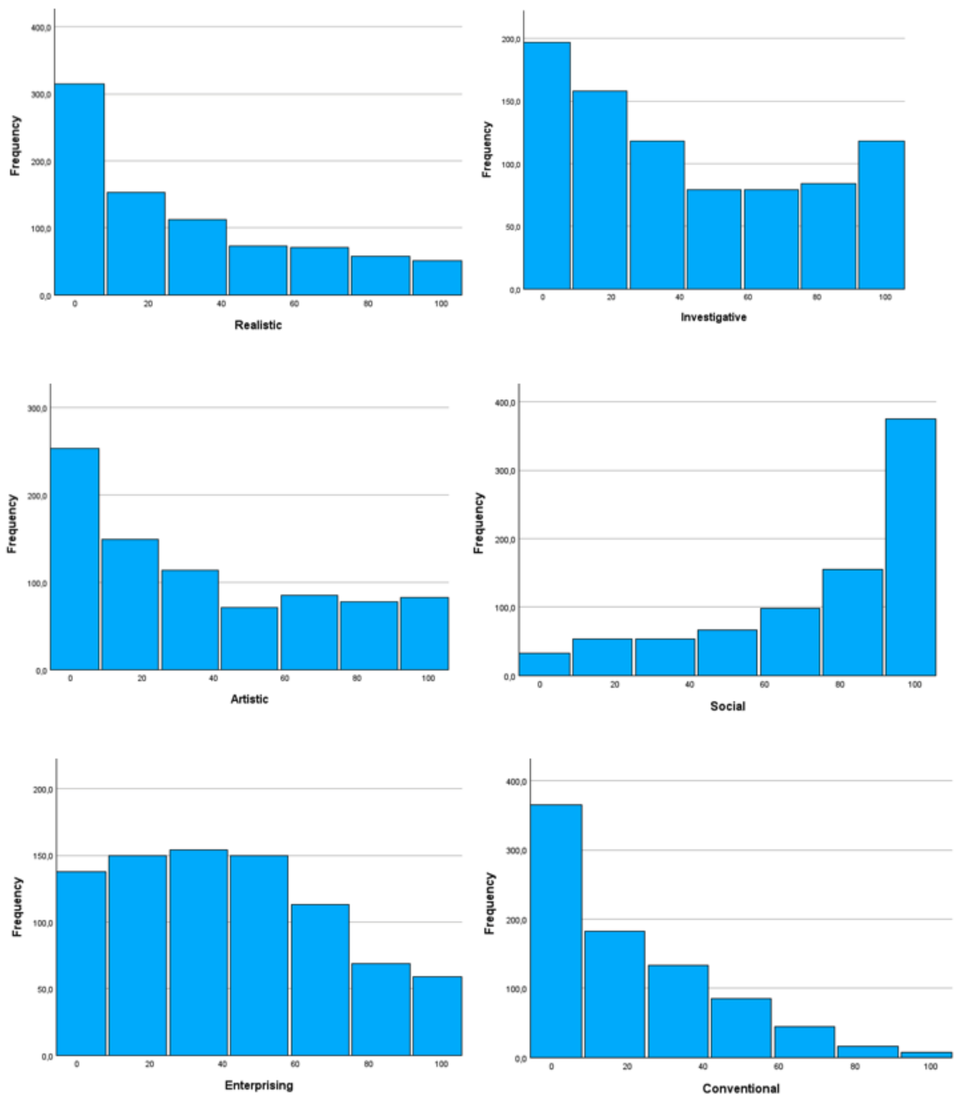


Table 1

Descriptive Statistics and Gender Comparisons for RIASEC Dimensions

	M *(SD)*	Boys: M *(SD)*	Girls: M *(SD)*	U	*p*	effect size *r*
Realistic	29.53 *(31.99)*	43.56 *(33.80)*	13.11 *(19.59)*	130,894.00	<.001	0.47
Investigative	41.52 (*35.54*)	45.30 *(35.73)*	37.02 *(34.79)*	96,650.50	.001	0.11
Artistic	36.37 *(34.54)*	20.33 *(24.79)*	55.26 *(34.97)*	37,988.00	<.001	0.49
Social	75.57 *(29.94)*	67.11 *(32.14)*	85.61 *(23.55)*	55,185.00	<.001	0.32
Enterprising	41.20 *(29.93)*	48.30 *(29.53)*	32.81 *(28.31)*	111,121.50	<.001	0.26
Conventional	20.11 *(23.39)*	22.15 *(23.69)*	17.76 *(22.90)*	95,242.0	.003	0.10

A CFA with the RIASEC-items serving as indicators for their respective latent interest dimensions (Realistic, Investigative, Artistic, Social, Enterprising, Conventional) confirmed an acceptable fit for the data: $\chi^2(579)$: 1,624.531; RMSEA: .047, 90% CI [.044, .049]; SRMR: .088; CFI: .965; TLI: .962. All factor loadings were significant, ranging from 0.549 to 0.960, indicating that observed variables reliably represent the underlying latent factors. The standardized factor loadings are presented in Appendix 4C.

Holland’s predicted pattern—where adjacent dimensions correlate higher than alternate dimensions, and opposing dimensions show the lowest correlation—was not confirmed (Table 2). This finding is further supported by the RTOR analysis[29], which yielded a correspondence index (CI) of .32, with a p-value of .17.

Table 2

Correlations Between the RIASEC Dimensions

	1.	2.	3.	4.	5.	6.
1. Realistic	1					
2. Investigative	.03	1				
3. Artistic	$-.20^{**}$	.02	1			
4. Social	$-.15^{**}$	$.10^{**}$	$.38^{**}$	1		
5. Enterprising	$.15^{**}$	$.16^{**}$	$-.12^{**}$	$.08^{*}$	1	
6. Conventional	$.16^{**}$	$.27^{**}$	-.02	$.14^{**}$	$.47^{**}$	1

Note. $^{*}p<.05$; $^{**}p < .01$.

Sex Differences

Mann-Whitney U tests revealed significant differences between boys and girls across all RIASEC interest dimensions, even after applying the Bonferroni-adjusted significance threshold ($p < .0083$; Table 1). Boys demonstrated stronger interests in the Realistic, Investigative, Enterprising, and Conventional dimensions, while girls exhibited stronger Artistic and Social interests. The largest differences were observed in the Realistic (effect size $r = .47$) and Artistic (effect size $r = .49$) dimensions.

The DIF analysis identified 13 items with (marginally) significant DIF (Table 3). Importantly, for all dimensions (except the Conventional dimension), there is an equal number of items that favor boys and girls. These results indicate that although there is sex bias in some of the interest items, this bias does not systematically affect the interest scale

[29]Holland's theory predicts the pattern of correlations among adjacent, alternate, and opposing scales. The RTOR analysis assesses the extent to which these predictions are met, expressed as a correspondence index (CI) ranging from -1 (no predictions confirmed) to +1 (all predictions confirmed). Benchmarks for interpreting the CI are .70 for U.S. samples and .48 for international contexts (Rounds & Tracey, 1996).

scores of a specific group. The DIF analysis results for all items are presented in Appendix 4D.

Table 3

Items Displaying DIF

Item n°	Mantel-Haenszel χ^2	*p*	log odds ratio effect size (DeltaMH)
R4	4.32	.038	1.22 (B)
R6	14.22	<.001	-2.50 (C)
I1	17.85	<.001	-1.94 (C)
I2	5.16	.023	1.21 (B)
A1	14.18	<.001	1.89 (C)
A6	4.43	.035	-1.15 (B)
S4	11.08	<.001	-2.10 (C)
S6	7.45	.006	1.61 (C)
E1	3.63	.057	0.88 (A)
E2	3.63	.057	0.92 (A)
E4	4.33	.037	-1.00 (B)
E6	3.80	.051	-1.08 (B)
C5	10.46	.001	1.70 (C)

Note. Effect size code A = negligible effect; B = moderate effect; C = large effect (Magis et al., 2010).

PE Interest Fit

The mean PE interest fit was $M = .57$ ($SD = .32$). Comparisons between boys and girls revealed a significant difference, as an independent samples t-test with sex as grouping variable and PE interest fit as dependent variable showed that girls ($M = .61$, $SD = .30$) had a higher PE interest fit than boys ($M = .54$, $SD = .34$): $t(567) = -2.61$, $p = .009$, *Cohen's d* = -.22.

No significant association was found between PE interest fit and GPA at T3 (r = .02, p = .57). When analyzed separately, this lack of association held true for both boys (r = -.02, p = .73) and girls (r = .03, p = .62). Similarly, analyses using within-school standardized GPA scores yielded consistent findings: r_{overall} = .02, p_{overall} = .67; r_{boys} = -.04, p_{boys} = .49; r_{girls} = .04, p_{girls} = .53

Discussion

The present study aimed to investigate PE interest fit in secondary education. While vocational interests, and specifically PE interest fit, have long been recognized as critical factors influencing academic and career outcomes, existing research has predominantly focused on adults or students in higher education. This focus leaves early adolescents largely understudied, despite the significant educational and career-related decisions early adolescents face during this formative period (Seghers et al., 2019). Given the proven value of PE interest fit in career education for older populations—where it has been shown to encourage exploration and self-reflection (Hanna & Rounds, 2020; Schelfhout et al., 2021b)—it is important to explore its potential in secondary education to help guide early adolescents in making informed and interest-driven choices.

The present study explored the fit between vocational interests and educational choices, potential sex differences, and the relation with academic achievement during early adolescence, using the SIMON-jr. The discussion will first address the validity of the SIMON-jr, followed by an analysis of the proposed hypotheses regarding PE interest fit.

SIMON-jr

The validity of the SIMON-jr was evaluated across three key aspects: internal consistency, internal structure, and sex differences. Regarding internal consistency, five dimensions (Realistic, Investigative, Artistic, Social, and Enterprising) exhibited reliability levels appropriate for less critical individual-level decisions (Evers et al., 2010), supporting their use in study counseling practices. The Conventional interest dimension showed comparatively lower internal consistency ($\alpha = .62$), though it still met the threshold for group-level research and fell within the reliability range observed in similar instruments, such as the Flemish ICA-R (De Vroede, 2015; Lavrijsen et al., 2021, p.3). While previous studies (Pässler & Hell, 2020; Lavrijsen et al., 2021) have also reported lower interest in the Conventional dimension compared to other RIASEC dimensions, the current study shows a more pronounced difference. This finding may reflect genuinely lower interest in the Conventional dimension within our sample, but could also suggest that the items representing this dimension were less clear or relevant for the population. Given these considerations, we suggest further research into the Conventional interest dimension to enhance understanding of both Conventional interests in early adolescents and SIMON-jr's ability to capture this dimension.

With respect to the internal structure, the CFA results demonstrated that the SIMON-jr aligns with the RIASEC framework. All factor loadings were significant, confirming that the SIMON-jr items reliably reflect the underlying RIASEC dimensions. The data did not support a circular structure, which aligns with prior research suggesting that the development of a circumplex relationship among interest scales typically emerges between seventh and eighth grade (ages 12-14) and becomes evident in eighth or ninth grade (around age 14-15; Tracey, 2002; Tracey & Rounds, 1993). Additionally, recent literature in career counseling and vocational interests increasingly recognizes the complexity of interest structures, particularly in early adolescence, suggesting that individuals in this age group may hold multiple vocational interests simultaneously, rather than adhering to the bipolarity principle of the RIASEC model (Slot et al., 2021; Tay et al. 2011).

Our findings concerning sex differences in vocational interests align with commonly observed patterns in early adolescence. Specifically, boys reported higher Realistic interests, while girls exhibited stronger Social and Artistic interests, consistent with prior research (Maurice & Bäumer, 2015; Pässler & Hell, 2020). In contrast to commonly observed trends in the Conventional interest dimension (Pässler & Hell, 2020; Su et al., 2009; Tracey, 2002), we found that boys showed stronger Conventional interests than girls. However, this result is not entirely unexpected as previous research within the same age group has also identified a reversed trend (Gfrörer et al., 2022). Furthermore, in line with previous findings in both early and late adolescent samples (Maurice & Bäumer, 2015; Su et al., 2009; Tracey, 2002), boys displayed stronger Investigative interests.

These findings regarding sex differences also highlight the importance of considering sex fairness in the SIMON-jr. DIF analysis showed bias in some individual items, though this bias was not systematically in favor of either boys or girls on any scale. These results suggest that while minor item-level differences exist, they do not undermine the overall fairness of the RIASEC scales.

In summary, our results support the reliability and validity of SIMON-jr for assessing vocational interests in early adolescents. The instrument proves a reliable and fair tool, that produces valid interest measures, which also make the tool a valuable resource for guiding students towards informed educational choices.

PE Interest Fit

Regarding the PE interest fit, our findings suggest that students generally select secondary education programs that fit reasonably well with their interests, with a mean fit of .57, supporting Hypothesis 1. However, this value is lower than the mean fit of .70 typically observed in higher education (Schelfhout et al., 2019). This difference aligns with existing literature, which indicates that early adolescents are often still in the process of developing their vocational interests (Slot et al., 2021; Sung et al., 2017); as they age, their vocational interests tend to become more differentiated and specific (Tracey, 2002), resulting in greater fit between interests and educational choices. The observation that girls demonstrated a stronger PE interest fit than boys (consistent with Hypothesis 3) further supports this interpretation, as previous research indicates that girls tend to exhibit more differentiated interest profiles in early adolescence (Slot et al., 2021; Sung et al., 2017). Another factor that may contribute to the lower PE fit in secondary education compared to higher education is the broader, more general nature of secondary education programs, particularly in academic tracks. Future research could investigate this explanation further by comparing the PE interest fit between vocational education programs, which typically offer more specialized curricula designed to prepare students for specific professions, and more theoretical programs, which, despite some unique characteristics, provide a broader and less specialized curriculum. To explore this explanation, we conducted an independent samples t-test using vocational education at T3 (yes or no) as the grouping variable and PE interest fit as the dependent variable. The analysis revealed that students in vocational education exhibited slightly higher PE interest fit ($M = .62$, $SD = .28$) than those in more theoretical tracks ($M = .58$, $SD = .32$). However, this difference was not statistically significant, $t(507) = -1.04$, $p = .30$, Cohen's d = -.12. Notably, there was a substantial disparity in sample sizes between the two groups, with only 93 students in vocational education compared to 416 in more theoretical tracks. Therefore, further research is recommended to more comprehensively examine the relationship between program type and PE interest fit.

Concerning the relationship between PE interest fit and academic achievement, our findings did not support Hypothesis 2, which predicted a positive relationship. Instead, no significant relationship was observed, nor when analyzing boys and girls separately. This result aligns with the findings of de Vries et al. (2024), whose meta-analysis reported only a small positive association between PE interest fit and academic achievement in higher education ($\rho = 0.10$, 95% CI [0.05, 0.14]). De Vries et al. (2024) attributed this weak relationship to the larger within-program variation in RIASEC profiles in educational settings compared to occupational contexts. Educational programs often prepare students for a broad range of careers rather than targeting a specific vocation, leading to greater diversity in program environments and smaller correlations with academic outcomes. This explanation is likely even more pronounced in secondary education, where study programs tend to be less specialized and more general compared to higher education, potentially accounting for the absence of a relationship in our study.

Two additional factors may further contribute to the absence of a relationship with academic achievement. First, most students in our sample exhibited a relatively high degree of interest fit with their chosen program. When variation in interest fit is limited, its potential impact on academic achievement may be less detectable (Tracey et al., 2012).

Second, the absence of a correlation may also be influenced by how academic achievement was operationalized in this study. Academic achievement was measured using GPA, which reflects performance across a broad curriculum encompassing general subjects such as mathematics, language, history, and ICT, regardless of the chosen program. These general subjects often do not align with students' specific vocational interests and are typically excluded from higher education curricula unless they are directly related to the chosen field of study. The principle of correspondence (Ajzen & Fishbein, 1977) suggests that predictive accuracy improves when the level of generality or specificity in both the predictor and outcome is aligned (Irving & Smith, 2020). Applying this principle, the general nature of GPA makes it a potentially less suitable measure for investigating the effect of interest fit on academic achievement, as it fails to directly capture students' vocational preferences. Future research could address these limitations by examining performance in program-specific subjects (e.g., mechanics, Latin, biology, physics, economics) that align more closely with students' vocational interests and chosen programs. Such an approach may provide a clearer understanding of the relationship between PE interest fit and academic outcomes in secondary education.

Practical Implications

Although vocational interests are still developing during early adolescence, students are already required to make educational and career-related decisions (Seghers et al., 2019). Choosing a study path is a complex process, and both students and parents often seek guidance and information. Exploration tools like SIMON-jr can play a valuable role in this process. By assessing the fit between personal interests and available study programs, SIMON-jr encourages reflection and exploration (Holland, 1997) and may highlight study paths that might otherwise be overlooked, either because they are not well-known to students and parents or are not initially recognized as a good fit. As such, SIMON-jr helps students make more informed choices by providing a clearer understanding of their interests and the educational options available to them. Moreover, this approach complements the often subjective recommendations of teachers by adding valuable objectivity (Seghers et al., 2019). Thus, while there is currently no evidence that SIMON-jr provides insight into students' chances of success within their chosen programs, the tool remains a valuable resource for guiding study choices and pathway selection among early adolescents.

Limitations and Suggestions for Future Research

A first limitation concerns the developmental stage of the participants. Early adolescence is a formative period for vocational interest development, characterized by the ongoing crystallization and differentiation of interests (Slot et al., 2021). Since students' vocational interests are likely still evolving at this stage, longitudinal studies are needed to explore how vocational interests and interest fit develop over time, as well as how the SIMON-jr captures these evolving interests and predicts educational and career outcomes.

Another limitation is the composition of the sample, which predominantly consisted of students who were interested in their chosen study program. Only 8% of students reported a lack of interest when asked directly[30], “Do you find your study program interesting?”, and the relatively high levels of PE interest fit across the sample reflect this. While this supports the validity of SIMON-jr in capturing interest fit, the restricted range of interest fit may have made it more difficult to detect significant outcomes. Future research could address this limitation by actively seeking a more diverse sample, particularly by focusing on smaller samples of targeted students who have switched programs or dropped out. Investigating the interest fit of these students could provide valuable insights into the consequences of misaligned interests, such as disengagement, academic difficulties, or dissatisfaction with their program. This approach would contribute to a more nuanced understanding of how PE interest fit influences educational choices and student success.

Conclusion

The present study supports the reliability and validity of SIMON-jr as an equitable tool for assessing vocational interests in early adolescents. Our findings indicate that secondary students tend to select education programs that align with their interests, with girls demonstrating a greater interest fit than boys, likely due to more differentiated interest profiles at this developmental stage. No significant relationship was found between interest fit and academic achievement. These results highlight the potential of career education tools that assess PE interest fit in secondary education, as they may encourage exploration and reflection on personal interests and study options, helping students make informed academic and career decisions.

[30] This question was included in the final data collection.

References

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin, 84,* 888-918, https://doi.org/10.1037/0033-2909.84.5.888

Allen J., & Robbins S. (2010). Effects of interest–major congruence, motivation, and academic performance on timely degree attainment. *Journal of Counseling Psychology, 57,* 23–35. https://doi.org/10.1037/a0017267 PMID: 21133558

Assouline, M., & Meir, E. I. (1987). Meta-analysis of the relationship between congruence and well-being measures. *Journal of Vocational Behavior, 31,* 319–332. https://doi.org/10.1016/0001-8791(87)90046-7

Chartrand, J., & Walsh, W. B. (1999). What should we expect from congruence? *Journal of Vocational Behavior, 55,* 136–146. https://doi.org/10.1006/jvbe.1999.1703

de Vries, N., Meeter, M., & Huizinga, M. (2024). Does interest fit between student and study program lead to better outcomes? A meta-analysis of vocational interest congruence as predictor for academic success. *Educational Research Review, 44,* 100619. https://doi.org/10.1016/j.edurev.2024.100619

Department of Education and Training. (2023). *Voltijds gewoon secundair onderwijs: algemeen overzicht* [Full-time regular secondary education: General overview] [Dataset]. Statistical yearbook of Flemish education 2023-2024. Flemish Government. https://onderwijs.vlaanderen.be/nl/onderwijsstatistieken/statistisch-jaarboek/statistisch-jaarboek-van-het-vlaams-onderwijs-2023-2024

De Vroede, T. (2015). *So you think you can choose. Interessemeting in de eerste jaren van het secundair onderwijs benaderd vanuit de Holland RIASOC-typologie. [So you think you can choose. Interest measurement in the first grades of secondary education* *according to the Holland-typology].* Unpublished master dissertation. Leuven: KU Leuven.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Review System for Evaluating Test Quality.* Boom Uitgevers Amsterdam. Retrieved from https://psynip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf

Flemish government. (n.d.). *Onderwijs in Vlaanderen* [Education in Flanders]. Onderwijs.vlaanderen.be. https://onderwijs.vlaanderen.be/

Fouad, N. A., & Mohler, C. J. (2004). Cultural validity of Holland's theory and the Strong Interest Inventory for five racial/ethnic groups. *Journal of Career Assessment, 12,* 423– 439. http://dx.doi.org/10.1177/10690 72704267736

Fonteyne, L., Wille, B., Duyck, W., De Fruyt, F. (2017). Exploring vocational and academic fields of study: development and validation of the Flemish SIMON Interest Inventory (SIMON-I). *International Journal for Educational and Vocational Guidance, 17,* 233-262. https://doi.org/10.1007/s10775-016-9327-9.

Gfrörer, T., Stoll, G., Rieger, S., Trautwein, U., & Nagengast, B. (2022). The development of vocational interests in early adolescence: Stability, change, and state-trait components. *European Journal of Personality, 36,* 942. https://doi.org/10.1177/08902070211035630

Goosen, K., Boone, S., Dehertogh, B., Kavadias, D., Mahieu, P., Van Avermaet, P., … Van Kerckhove, C. (2017). *Is dat iets voor mij juf? Leerlingen versterken in het keuzeproces van basis naar secundair.* Leuven: LannooCampus.

Granillo-Velasquez, K. E., Hoff, K. A., Hanna, A., Oswald, F. L., & Morris, M. L. (2024). Comparing methods of measuring interest fit: A large prediction study with career choice satisfaction. *International Journal of Selection and Assessment,* 1–13. https://doi.org/10.1111/ijsa.12506

Hanna, A., Briley, D., Einarsdóttir, S., Hoff, K., Rounds, J. (2021). Fit gets better: A longitudinal study of changes in interest fit in educational and work environments. *European Journal of Personality, 35,* 557-580. https://doi.org/10.1177/08902070211014022

Hanna, A., & Rounds, J. (2020). How accurate are interest inventories? A quantitative review of career choice hit rates. *Psychological Bulletin, 146,* 765–796. https://doi.org/10.1037/bul0000269

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review, 1,* 69–82. https://doi.org/10.1016/j.edurev.2006.09.001.

Hoff, K. A., Song, Q. C., Wee, C. J., Phan, W. M. J., & Rounds, J. (2020). Interest fit and job satisfaction: A systematic review and meta-analysis. *Journal of Vocational Behavior, 123,* Article 103503. https://doi.org/10.1016/j.jvb.2020.103503

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL, US: Psychological Assessment Resources.

Hubert, L., & Arabie, P. (1987). Evaluating order hypotheses within proximity matrices. *Psychological Bulletin, 102,* 172-178. https://doi.org/10.1037/0033-2909.102.1.172

Hughes, D., Mann, A., Barnes, S.-A., Baldauf, B., & McKeown, R. (2016). *Careers education: International literature review.* Warwick Institute for Employer Research and Education and Employers Research. https://www.educationandemployers.org/wp-content/uploads/2016/07/Careers-review.pdf

Hyland, W. E., Hoff, K. A., & Rounds, J. (2022). Interest–ability profiles: An integrative approach to knowledge acquisition. *Journal of Intelligence, 10,* 43. https://doi.org/10.3390/jintelligence10030043

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29.0). Armonk, NY: IBM Corp

Irving, L. H., & Smith, C. T. (2020). Measure what you are trying to predict: Applying the correspondence principle to the Implicit Association Test. *Journal of Experimental Social Psychology, 86,* 103898. https://doi.org/10.1016/j.jesp.2019.103898

Kuijpers, M., & Scheerens, J. (2006). Career competencies for the modern career. Journal of *Career Development, 32,* 303–319. https://doi.org/10.1177/ 0894845305283006

Lasselle, L., Schelfhout, S., Fonteyne, L., Kirby, G., Smith, I., & Duyck, W. (2021). An examination of gender imbalance in Scottish adolescents' vocational interests. *PLOS ONE, 16*. https://doi.org/10.1371/journal.pone.0257723

Lavrijsen, J., Terence J. G. Tracey, T. J. G., Verachtert, P., De Vroede, T., Soenens, B. & Verschueren, K. (2021). Understanding school subject preferences: The role of trait interests, cognitive abilities and perceived engaging teaching. *Personality and Individual Differences, 174,* 110685. https://doi.org/10.1016/j.paid.2021.110685

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*,79-122. https://doi.org/10.1006/jvbe.1994.1027

Magis, D., Beland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods, 42,* 847–862. https://doi.org/ 10.3758/BRM.42.3.847

Maurice, J. V., & Bäumer, T. (2015). Entwicklung allgemeiner Interessenorientierungen beim Übergang von der Grundschule in den Sekundarbereich. In C. Tarnai, & F. G.

Hartmann (Eds.). *Berufliche Interessen: Beiträge zur Theorie von J. L. Holland* (pp. 64–85). Waxmann.

McLarnon, M. J., Carswell, J. J., & Schneider, T. J. (2015). A case of mistaken identity? Latent profiles in vocational interests. *Journal of Career Assessment, 23,* 166–185. http://doi.org/10.1177/10690 72714523251

Nagy, G., Trautwein, U., & Lüdtke, O. (2010). The structure of vocational interests in Germany: Different methodologies, different conclusions. *Journal of Vocational Behavior, 76,* 153–169. http://dx.doi.org/10.1016/ j.jvb.2007.07.002

Nauta, M. M. (2010). The development, evolution, and status of Holland's theory of vocational personalities: Reflections and future directions for counseling psychology. *Journal of counseling psychology, 57,* 11-22. https://doi.org/10.1037/a0018213 PMID: 21133557

Nye, C. D., Butt, S. M., Bradburn, J., & Prasad, J. (2018). Interests as predictors of performance: An omitted and underappreciated variable. *Journal of Vocational Behavior, 108,* 178-189. https://doi.org/10.1016/j.jvb.2018.08.003

Nye, C. D., Prasad, J., Bradburn, J., & Elizondo, F. (2018). Improving the operationalization of interest congruence using polynomial regression. *Journal of Vocational Behavior, 104,* 154-169. https://doi.org/10.1016/j.jvb.2017.10.012

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7,* 384–403. https://doi.org/10.1177/1745691612449021

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior, 98,* 138–151. https://doi.org/10.1016/j.jvb.2016.11.002

Pässler, K., & Hell, B. (2020). Stability and change in vocational interests from late childhood to early adolescence. *Journal of Vocational Behavior, 121,* 103462. https://doi.org/10.1016/j.jvb.2020.103462

Perera, H. N., & McIlveen, P. (2018). Vocational interest profiles: Profile replicability and relations with the STEM major choice and the Big-Five. *Journal of Vocational Behavior, 106,* 84 –100. http://doi.org/10.1016/ j.jvb.2017.11.012

Pustjens, H., Van de Gaer, E., Van Damme, J., & Onghena, P. (2008). Curriculum choice and success in the first two grades of secondary education: Students, classes, or schools? *School Effectiveness and School Improvement, 19,* 155-182. https://doi.org/10.1080/09243450802047923

R Core Team. (2024). *R: A language and environment for statistical computing* (version 4.3.3) [Software]. Vienna, Austria. Retrieved from http://www.R-project.org/

Rosseel, Y. (2012). Lavaan: An R package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1-36. https://doi.org/10.18637/jss.v048.i02

Rounds, J., Hoff, K. A., & Lewis, P. (Eds.). (2021). *O*NET® Interest Profiler manual.* The National Center for O*NET Development. https://www.onetcenter.org/dl_files/IP_Manual.pdf

Rounds, J., & Su, R. (2014). The nature and power of interests. *Current Directions in Psychological Science, 23,* 98–103. https://doi.org/10.1177/0963721414522812.

Rounds, J., & Tracey, T. J. G. (1996). Cross-cultural structural equivalence of RIASEC models and measures. *Journal of Counseling Psychology, 43,* 310. https://doi.org/10.1037/0022-0167.43.3.310

Rounds, J., Walker, C. M., Day, S. X., Hubert, L., Lewis, P., & Rivkin, D. (1999). *O*NET Interest Profiler: Reliability, validity, and self-scoring.* https://www.onetcenter.org/reports/IP_RVS.html

Schelfhout, S., Bassleer, M., Wille, B., Van Cauwenberghe, S., Dutry, M., Fonteyne, L., … Duyck, W. (2022). Regressed person-environment interest fit: Validating polynomial regression for a specific environment. *Journal of Vocational Behavior, 136.* https://doi.org/10.1016/j.jvb.2022.103748

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2019). The effects of vocational interest on study results: Student Person – Environment fit and program interest diversity. *PLoS One, 14*, e0214618, https://doi.org/10.1371/journal.pone.0214618

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2021b). From interest assessment to study orientation: An empirical advice set engine. *The Journal of Experimental Education, 89,* 169–195. https://doi.org/10.1080/00220973.2019.1605327

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., Derous, E., De Fruyt, F., & Duyck, W. (2021a). How interest fit relates to STEM study choice: Female students fit their choices better. *Journal of Vocational Behavior, 129,* 103614. https://doi.org/10.1016/j.jvb.2021.103614

Seghers, M., Boone, S., & Van Avermaet, P. (2019). Classed patterns in the course and outcome of parent-teacher interactions regarding educational decision-making. *Educational Review, 73,* 417-435. https://doi.org/10.1080/00131911.2019.1662771

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010). Polynomial regression with response surface analysis: A powerful approach for examining moderation and overcoming limitations of difference scores. *Journal of Business and Psychology, 25,* 543–554. https://doi.org/10.1007/s10869-010-9183-4

Slot, E. M., Bronkhorst, L. H., Akkerman, S. F., & Wubbels, T. (2021). Vocational interest profiles in secondary school: Accounting for multiplicity and exploring associations with future-oriented choices. *Journal of Educational Psychology, 113,* 1059-1071. https://doi.org/10.1037/edu0000475

Su, R. (2020). The three faces of interests: An integrative review of interest research in vocational, organizational, and educational psychology. *Journal of Vocational Behavior, 116,* 103240. https://doi.org/10.1016/j.jvb.2018.10.016.

Su R., Rounds J., Armstrong P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological bulletin, 135,* 859-884. https://doi.org/10.1037/a0017364

Sung, Y. T., Cheng, Y. W., & Hsueh, J. H. (2017). Identifying the career-interest profiles of junior-high-school students through latent profile analysis. *The Journal of Psychology, 151,* 229 –246. http://doi .org/10.1080/00223980.2016.1261076

Tay, L., Su, R., & Rounds, J. (2011). People—Things and data—Ideas: Bipolar dimensions? *Journal of Counseling Psychology, 58,* 424 – 440. http://doi.org/10.1037/a0023488

Tracey, T. J. G. (2002). Development of Interests and Competency Beliefs: A 1-Year Longitudinal Study of Fifth- to Eighth-Grade Students Using the ICA–R and Structural Equation Modeling. *Journal of Counseling Psychology, 49* 148-163. https://doi.org/10.1037//0022-0167.49.2.148

Tracey, T. J. G., & Rounds, J. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin, 113*, 229-246. https://doi.org/10.1037/0033-2909.113.2.229

Tracey, T. J. G., & Rounds J. (1996). The spherical representation of vocational interests. *Journal of Vocational Behavior, 48,* 3–41. https://doi.org/10.1006/jvbe.1996.0002

Tracey, T.J.G., Allen, J., & Robbins, S.B. (2012). Moderation of the relation between personen-vironment congruence and academic success: environmental constraint, personal flexibility and method. *Journal of Vocational Behavior, 80,* 38-49. https://doi:10.1016/j.jvb.2011.03.005

Tracey, T. J., & Ward, C. C. (1998). The structure of children's interests and competence perceptions. *Journal of Counseling Psychology, 45*, 290-303. https://doi.org/ 10.1037/0022-0167.45.3.290

Van Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: The validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology, 96,* 13–33. https://doi.org/10.1037/a0021193.

Wiegand, J. P., Drasgow, F., & Rounds, J. (2021). Misfit matters: A re-examination of interest fit and job satisfaction. *Journal of Vocational Behavior, 125,* 103524. https://doi.org/10.1016/j.jvb.2020.103524

Wille, E., Stoll, G., Gfrörer, T., Cambria, J., Nagengast, B., & Trautwein, U. (2020). It takes two: expectancy-value constructs and vocational interests jointly predict STEM major choices. *Contemporary Educational Psychology, 61,* 101858. https://doi.org/10.1016/j.cedpsych.2020.101858

Xia, Y., & Yang, Y. Y. (2018). The influence of number of categories and threshold values on fit indices in structural equation modeling with ordered categorical data. *Multivariate Behavioral Research, 53,* 731-755. https://doi.org/ 10.1080/00273171.2018.1480346

Xu, H., & Li, H. (2020). Operationalize interest congruence: A comparative examination of four approaches. *Journal of Career Assessment, 28,* 571–588. https://doi.org/10.1177/1069072720909825

Xu, H., & Tracey, T. J. G. (2016). Stability and change in interests: A longitudinal examination of grades 7 through college. *Journal of Vocational Behavior, 93,* 129–138. https://doi.org/10.1016/j.jvb.2016.02.002.

Chapter 5.

Interest and Person-Environment Interest Fit Development in Early Adolescence

Submitted to Journal of Vocational Behavior, April 2025.

Merel Dutry, Stijn Schelfhout, Eva Derous, Wouter Duyck, and Nicolas Dirix.

Abstract

This study explores the longitudinal development of Person-Environment (PE) interest fit during early adolescence within the Flemish educational system. While theories of PE fit suggest that interest fit improves over time due to socialization and selection processes, less is known about how these processes unfold during early adolescence, a stage marked by evolving interests and key educational choices. By focusing on this formative period, the present longitudinal study offers a unique perspective on the dynamics of PE interest fit, with implications for both theoretical models and practical applications in education and career counseling. We examine how vocational interests and PE interest fit evolve from grade 7 (ages 12–13) to grade 9 (ages 14–15), while also considering potential gender differences, through multilevel models. The results demonstrate that students already exhibit a substantial fit with their grade 9 program choice at the start of secondary education. While students' interests evolve during adolescence, the fit with their chosen study program remains stable over time. Instead, students—particularly boys—tend to progressively lose interest in the options they have excluded, reflecting a narrowing of focus. Moreover, for both boys and girls, the difference between the fit with chosen and unchosen programs becomes increasingly pronounced over time, primarily driven by the decline in fit with unchosen options. The findings highlight the importance of studying the factors influencing the development of PE interest fit during early adolescence and provide practical insights for using interest assessments in secondary education.

Introduction

Young adolescents often face the task of making educational choices that are expected to fit with their interests (Seghers et al., 2019). However, as early adolescence represents a formative period for the development and stabilization of interests (Hoff et al., 2018), aligning these choices with emerging and sometimes dynamic preferences can be a significant challenge (Slot et al., 2021). To support adolescents in navigating these complex decisions, researchers and practitioners must understand how interests and person-environment (PE) interest fit—the compatibility between a person's attributes and those of a (study) environment (Hanna et al., 2021)—evolve during this period. While studies have documented patterns of both stability and change in vocational interest dimensions during early adolescence (e.g., Gfrörer et al., 2022; Pässler & Hell, 2020; Tracey, 2002; Tracey & Sodano, 2008), the development of the fit of those interests with the environment in which those adolescents function (i.e. school) remains largely unexplored.

To address this gap in literature, the present study examines the longitudinal development of PE interest fit during early adolescence within the Flemish educational system. We also examine patterns of stability and change in singular vocational interest dimensions to explore how these changes shape the development of PE interest fit. More specifically, we investigate the evolution of fit towards which program is chosen, and which are not, while also controlling for gender differences.

RIASEC Theory

Holland's (1997) theory of vocational interests is the most extensively used framework for interest classification (Hoff et al., 2018). According to Holland (1997), the interest profiles of both individuals and environments can be described using six primary dimensions—Realistic (R), Investigative (I), Artistic (A), Social (S), Enterprising (E), and Conventional (C)—collectively referred to as RIASEC. These dimensions represent trait-like preferences that encompass values, abilities, and favored activities, offering insight into both individual and environmental characteristics. Holland's framework has been rigorously validated over decades and forms the basis for numerous interest inventories (Lasselle et al., 2021; Nauta, 2010), including SIMON-jr (Dutry et al., *In Preparation*) used in this study to assess the vocational interests of early adolescents and define their RIASEC profiles.

A RIASEC test allocates a score to each RIASEC interest dimension, which together constitute the individual's RIASEC profile (Lasselle et al., 2021). The incumbent method is a typical method to construct RIASEC profiles for environments (Allen & Robbins, 2010). Building on the common principle that individuals shape their own environment (Schneider, 1987), this method uses average RIASEC scores of students within a certain study program to represent the program's RIASEC profile (Allen & Robbins, 2010; Schelfhout et al., 2022). By using students as incumbents, this method creates empirically grounded environment profiles that reflect the collective interests of students in a specific environment.

Commensurate RIASEC measurement of both persons (P) and environments (E) enables the assessment of PE fit, linking personal interests with the characteristics of their environment (Nye et al., 2012; Schelfhout et al., 2021a). PE fit theory posits that individuals are inclined toward environments that fit with their interests and that they are more likely to excel and experience satisfaction when their academic or professional settings match their preferences (Hoff et al., 2020; Holland, 1997; Lent et al., 1994; Tracey et al., 2012). Supporting this theoretical perspective, recent meta-analyses reveal positive relations between PE fit and key outcomes like performance and satisfaction in the workplace (Hoff et al., 2020; Nye et al., 2017; Van Iddekinge et al., 2011), and academic achievement, persistence, and overall contentment in educational settings (de Vries et al., 2024; Nye et al., 2012). When selecting a method to assess PE interest fit, both predictive capability and interpretability are crucial considerations (Granillo-Velasquez et al., 2024). Indeed, the observed and predicted relationships often differ based on the type of interest measurement and the method used to operationalize PE fit (Xu & Li, 2020). Profile correlation is one pattern-based similarity measure that uses Pearson's correlation to compare the RIASEC profiles of students and study programs (Schelfhout et al., 2019). The resulting correlation reflects how well a student's interests fit with the chosen study program, with higher correlations indicating a better fit. This method is one of the preferred methods in career education, where the goal is to recommend paths aligned with individual interests, and is applied in the current study to operationalize PE interest fit.

Vocational Interest Development in Early Adolescence

Early adolescence is a crucial period in personal development, marked by changes in school, peer relationships, and family dynamics (Salgado et al., 2021). During this time, adolescents gain greater autonomy from their parents (Keijsers & Poulin, 2013), which broadens their opportunities to explore new activities. As young adolescents become more capable of selecting environments and activities that reflect their evolving interests, they lay the foundation for their vocational interest profiles (Gfrörer et al., 2022; Tracey & Sodano, 2008). This formative period makes the study of vocational interest development during early adolescence especially compelling.

Research on vocational interest development often utilizes various metrics to examine patterns of stability and change (Tracey & Sodano, 2008; Xu & Tracey, 2016). Two commonly used approaches in this context are mean-level changes and test-retest correlations of interest assessments (Gfrörer et al., 2022). Mean-level changes reflect shifts in the overall intensity of interests, averaged across a sample, and are commonly referred to as normative changes, representing broad patterns of personality development that are applicable to the majority of individuals (Hoff et al., 2018; Roberts et al., 2006). In contrast, test-retest correlations serve as indicators of construct stability, capturing either the consistency of individuals' relative rank order within a group over time (interindividual interest stability) or the consistency of a person's RIASEC profile over time (intraindividual stability) (Ertl et al., 2023; Low & Rounds, 2005).

The meta-analysis of Low et al. (2005) demonstrates that vocational interests remain relatively stable from early adolescence to middle adulthood, with increasing coherence and differentiation during adolescence—patterns similar to those observed in personality trait development (Bleidorn et al., 2022). Specifically, Low et al. (2005) found that vocational interests exhibit moderate test-retest correlations in early adolescence ($r = .55$), which increase to relatively high levels in adulthood (up to $r = .83$). More recent studies further support this trend, shedding light on age-related changes in vocational interests. For instance, a longitudinal study tracking 3,786 German students from 5th to 8th grade (ages 10-14; Gfrörer et al., 2022) report moderate ($.41 < r < .64$) test-retest correlations and a significant increase in the stability of Realistic, Investigative, Social, and Conventional interest dimensions over time, along with an increasing proportion of trait variance in vocational interests as students matured. Similarly, a study involving 9,494 Icelanders aged 13 to 60 (Einarsdóttir & Rounds, 2020) found that younger participants (ages 13–14) exhibited less distinct and more mixed RIASEC patterns compared to the clearer, more differentiated structures seen in older age groups. Likewise, Pässler and Hell (2020) investigated vocational interest development in a longitudinal sample of Swiss children aged 10 to 12, reporting small to moderate stability levels.

This increasing stability observed in adolescents may be attributed to their engagement in a broader range of activities over time, allowing to clarify their vocational interests and build a stronger sense of identity (Gfrörer et al., 2022). These patterns of increasing stability align with the theoretical assumptions of Holland (1997) and the Trait Situation Interest Dynamic (TSID) model presented by Su et al. (2019). Holland (1997) proposed that interest stability emerges through cumulative learning, where repeated engagement in activities gradually reinforces and solidifies specific preferences. In parallel, the TSID model posits that vocational interests are shaped and refined through positive experiences or adjusted in response to negative ones.

Despite this overall stability, evidence suggests temporary mean-level declines in vocational interests during adolescence, within the generally upward trajectories observed over a longer time (Hoff et al., 2018; Pässler & Hell, 2020). The disruption hypothesis, commonly applied to personality traits, attributes such declines to the significant physiological, social, and normative changes occurring during adolescence (Hoff et al., 2018; Pässler & Hell, 2020; Soto & Tackett, 2015). These challenges may hinder the fulfillment of basic psychological needs—competence, autonomy, and relatedness—which are essential for fostering interest (Ryan & Deci, 2000), leading to a general decline in RIASEC scores. Hoff et al. (2018) indeed reported general mean-level declines in vocational interests between ages 11 and 14, particularly in the Conventional ($d = –0.30$) and Realistic ($d = –0.17$) dimensions, with a modest increase in Enterprising interests ($d = 0.16$). Similarly, Pässler and Hell (2020) also observed significant decreases in Realistic, Investigative, Artistic, and Social interests between ages 10 and 12. Gfrörer et al. (2022) further documented declines over the course of late childhood and early adolescence in Realistic, Investigative, Artistic, and Conventional interests, with effect sizes ranging from -0.44 to -0.01.

In addition to developmental disruptions, these temporary declines can also be understood through Gottfredson’s (1981) theory of circumscription and compromise. This theory proposes that through experiences and interactions, young people gradually eliminate occupations that do not align with their evolving self-concept. Following this "circumscription" phase, adolescents enter a phase of "compromise”, where they begin to recognize that some occupations they see as a good fit for their self-concept may not be realistically available. As a result, fewer occupations feel suitable over time, and this mismatch between self-perception and potential career options may lead to a reduction in the intensity of their vocational interests, contributing to the observed mean-level declines (Gfrörer et al., 2022; Tracey, 2002; Krapp, 2002).

Despite extensive theoretical and empirical work on the development of vocational interests during early adolescence, researchers have paid little attention to how these dynamics influence the PE fit of evolving interests with the environment. This is the focus of the present study.

PE Interest Fit Development in Early Adolescence

While evolving vocational interests are part of personality development, changes in singular interest dimensions can affect PE interest fit (Ertl et al., 2023). Holland's (1997) theory suggests that people seek environments that fit with their vocational interests. Within such environments, individuals are likely to further develop their interests to align with environmental demands, enhancing the fit over time. Consequently, students entering the same academic environment may exhibit convergent vocational interest development due to shared requirements, whereas those pursuing different fields may show divergent development patterns (Ertl et al., 2023; Holland, 1997).

Similarly, Schneider's (1987) Attraction-Selection-Attrition (ASA) framework illustrates the dynamic processes of PE fit, emphasizing the role of vocational interests. This framework proposes that individuals are attracted to, enter, and leave jobs based on the fit between their personal attributes and the characteristics of the environment. Over time, the ASA cycle facilitates movement toward better-fitting environments, reinforcing and amplifying personality traits within these settings (Schneider, 1987). This process suggests a gradual improvement in PE interest fit, driven by mechanisms like socialization and selection (Hanna et al., 2021; Holland, 1997; Schneider, 1995).

While empirical research supports the roles of selection (Hanna et al., 2021; Roberts & Robins, 2004; Wille & De Fruyt, 2014) and socialization (Ertl et al., 2023; Nye et al., 2021) in shaping PE fit during late adolescence and adulthood, less is known about how these processes unfold during early adolescence—a stage characterized by evolving interests and pivotal educational choices (Slot et al., 2021).

Recent research by Dutry et al. (*In Preparation*) investigated how students' vocational interests in grade 7 (ages 12–13) align with their study program choices by grade 9 (ages 14–15). The findings suggest that while students generally choose educational programs that reflect their interests in grade 7, the fit is weaker compared to what is typically observed in higher education (Schelfhout et al., 2019). Given that vocational interests continue to evolve during early adolescence, these changes may influence the stability and predictive power of PE interest fit over time. Therefore, studying the longitudinal patterns of PE interest fit during this formative period is crucial, as it offers valuable insights into the underlying developmental processes, and informs the improvement of academic counseling practices in secondary education.

Building on the seminal work of Holland (1997) and Schneider (1987), and the more recent findings of Hanna et al. (2021), which suggest that individuals gravitate toward environments aligning with their evolving interests, we propose that interest fit improves over time. Accordingly, we hypothesize:

Hypothesis 1 (H1): PE interest fit with the chosen study program in grade 9 (ages 14-15) improves progressively throughout middle school (ages 12-14).

Furthermore, Gottfredson's (1981) theory of circumscription and compromise offers valuable insights into why PE fit may decline for paths individuals do not choose. According to this theory, as individuals narrow their focus toward specific career paths, they may perceive less fit with unchosen alternatives. This narrowing process, which is crucial for aligning interests with realistic options, likely contributes to a reduction in PE fit with non-chosen study programs. Given that vocational interests become more defined during adolescence, we hypothesize:

Hypothesis 2 (H2): PE interest fit with unchosen study programs in grade 9 (ages 14-15) diminishes throughout middle school (ages 12-14).

While both hypotheses appear related, the second is neither redundant nor necessarily consistent with the first, as they address different aspects of how vocational interests interact with PE interest fit and may be independent of each other. H1 suggests that as students' interests evolve, they gravitate toward the program that best fits with their goals. In contrast, H2 is based on the concept of *compromise*, where students reduce their interest in alternative options as they refine their self-concept. These processes can operate independently: The fit with the chosen program may improve without directly affecting the fit with unchosen options, and vice versa.

Gender Differences

Gender Differences in Vocational Interests

Gender differences in vocational interests are consistently observed across adolescence and adulthood. Men typically score higher on Realistic and Investigative dimensions, while women tend to score higher on Artistic, Social, and Conventional interests (Su et al., 2009). These differences are more pronounced in younger adults compared to older ones (Morris, 2016; Su et al., 2009). Similar patterns emerge in late childhood and early adolescence, where boys show stronger interests in Realistic and Investigative dimensions, and girls exhibit greater preferences for Artistic and Social interests (Gfrörer et al., 2022; Tracey, 2002). During early adolescence, these gender differences often increase, particularly in Realistic and Social interests (Gfrörer et al., 2022; Hoff et al., 2018). However, by late adolescence, these gender differences tend to diminish.

Regarding stability patterns, meta-analytic evidence from two-wave studies indicates no gender differences in rank-order stabilities (Low et al., 2005). However, longitudinal research on high school students across three time points has produced mixed findings. Some studies demonstrate that female adolescents exhibit slightly higher rank-order and profile stabilities (Tracey et al., 2005; Xu & Tracey, 2016), while others indicate no gender differences (Tracey & Robbins, 2005).

Given the consistent evidence of gender differences in vocational interests across adolescence and adulthood, along with mixed findings for developmental patterns, gender serves as an important control variable in this study. Accounting for gender differences allows for a more nuanced understanding of the development of vocational interests, thereby enhancing the validity and robustness of the findings. Moreover, examining gender differences can help identify potential barriers or opportunities, supporting efforts to promote more equitable educational outcomes and career paths.

Gender Differences in PE Fit

Gender differences are also evident in profile differentiation during adolescence (Xu & Tracey, 2016), with girls demonstrating more differentiated RIASEC profiles compared to boys (Fouad & Mohler, 2004; Slot et al., 2021; Sung et al., 2017). Differentiated profiles are associated with a stronger fit between an individual's interests and their educational choices, whereas less differentiated profiles often correspond to more varied choices (Slot et al., 2021). This pattern is reflected in PE fit, with studies generally showing that female students tend to achieve higher PE interest fit with their study programs than their male counterparts (e.g., Allen & Robbins, 2010; Tracey et al., 2012; Schelfhout et al., 2021b). Based on these findings, we hypothesize:

Hypothesis 3 (H3): PE interest fit with the chosen study program in grade 9 (ages 14-15) will be higher for girls than for boys.

Present Study

The present study examines the development of PE interest fit during early adolescence within the Flemish educational system. Using a longitudinal sample of Flemish secondary students, with data collected annually over three consecutive school years, we investigate how vocational interests evolve from grade 7 (ages 12–13) to grade 9 (ages 14–15), and how these evolving interests affect PE fit with the chosen and unchosen study programs in grade 9. Grade 9 in Flanders marks a pivotal transition, as students choose a study program that serves as a preliminary pathway to future careers or higher education.

The present study addresses two key research questions. The first examines the development of vocational interests, focusing on both their stability and change over time. Specifically, we analyze rank-order stability, mean-level changes, RIASEC profile stability, and the differentiation of interest profiles, with particular attention to potential gender differences. The second research question investigates the development of PE interest fit during secondary education within the Flemish educational system, testing the three hypotheses outlined in the introduction.

Method

Participants

This study draws on data from the Talentcenter Study, a project designed to help secondary students make well-informed study choices by providing personalized feedback on their interests, talents, and matching study programs (https://talentcenter.vlaanderen/). During the tool's development, Flemish students were tracked over the first three years of secondary education (grade 7 to grade 9), with data collected annually across three waves. The sample includes 885 students who contributed to at least one wave of data collection. Of these, 10% reported speaking a home language other than Dutch, with 11% of responses unspecified.

Furthermore, 27% had parents who received a scholarship, though financial status was unspecified for 20% of participants. Regarding parental education, 9% of participants had one parent with primary education or less as highest level of schooling, and 5% had both parents with no education beyond primary school (with 12% unspecified). Participants pursued a variety of study programs starting from grade 9, including academic, vocational, and technical track, which allowed us to capture students from a broad range of educational backgrounds, rather than focusing solely on academic tracks.

During the first measurement wave (T1 from September-November 2021, grade 7), N_{T1} = 833 students from 23 different regular secondary schools across Flanders completed the vocational interest survey ($M_{\text{age\_T1}}$ = 12.02; $SD_{\text{age\_T1}}$ = .55), with 46% identifying as a girl. During the second wave (T2 from February-May 2023, grade 8), N_{T2} = 718 students participated ($M_{\text{age\_T2}}$ = 13.50; $SD_{\text{age\_T2}}$ = .62; 45% girls). During the third and final wave (T3 from January-February 2024, grade 9), N_{T3} = 545 students remained[31] in the study and completed the survey, now spread across 36 different schools ($M_{\text{age\_T3}}$ = 14.22; $SD_{\text{age\_T3}}$ = .50; 46% girls). Information about the chosen study program was unavailable for 5 of these 545 students, as they either changed schools or programs during the final measurement wave or were transferred to special education. Additionally, we received study program information for 66 students who were absent during the third and final wave of data collection. This supplementary data was included in the PE fit analyses.

[31]Dropout in our study occurred for several reasons: some students were absent on the day of data collection (which took place during the COVID-19 pandemic), while others transferred schools. Students who switched schools were contacted via email, as class participation was no longer possible. However, many did not respond to email requests, likely due to disinterest or outdated contact information. A small number of students voluntarily withdrew from the study, in accordance with their rights as outlined in the informed consent, without providing a reason. The transition from grade 8 to grade 9 presented additional challenges, as some students attended schools that only offered the first two years of secondary education, requiring them to switch schools to continue their education. This dispersion across multiple, non-participating schools, combined with outdated contact details, made it difficult to reach many students for the third and final wave of data collection.

Of the total sample, 478 students (48% girls) participated in all three waves, representing 54% of all students. Comparisons of demographic characteristics showed that the sample of students who completed all waves included a relatively higher proportion of girls and came from slightly more privileged backgrounds compared to those who did not participate in all three waves (Table 5A1). However, there were no significant differences in vocational interests at T1 between students who completed all waves versus those who did not: Realistic ($t(831) = 1.31$, $p = .189$), Investigative ($t(831) = -1.59$, $p = .113$), Artistic ($t(831) = -.20$, $p =. 845$), Social ($t(831) = -1.94$, $p = .053$), Enterprising ($t(831)= .88$, $p = .380$), and Conventional ($t(831) = .58$, $p = .559$). Based on these findings, we concluded that data for vocational interests is missing at random, and no imputation was performed for missing data.

The research project[32] encompassing the present study was approved by the Ethical Commission of the Faculty of Psychology and Educational Sciences of Ghent University (reference number 2021/59). Both the parents of the participants and the participants themselves gave their informed consent for participation, as well as for storage and use of the data by the researchers.

Materials

Vocational Interests

We assessed students' vocational interests using the SIMON-jr inventory (Dutry et al., *In Preparation*), which is based on Holland's (1997) RIASEC model. The inventory contains 36 items, with 6 items for each interest dimension. These items describe either activities (26 items) or occupations (10 items) linked to a specific interest dimension. Participants indicate whether they would enjoy the activity or occupation by responding "yes" or "no." For each interest dimension, the percentage of “yes” responses is computed and used to construct the

[32]Entitled: “Study orientation in secondary education”.

participant's RIASEC profile. Table 1 presents the internal consistency for each RIASEC dimension across the measurement waves, based on all available data. Cronbach's alpha values range from .62 to .88, with internal consistency generally improving from T1 to T3.

Table 1

Means, Standard Deviations and Internal Consistencies of the RIASEC Scales across Three Measurement Waves

	M	*SD*	Cronbach's alpha (α)
Realistic T1	29.53	31.99	.80
Realistic T2	24.51	32.24	.85
Realistic T3	24.31	32.79	.86
Investigative T1	41.52	35.54	.83
Investigative T2	39.39	36.70	.85
Investigative T3	39.24	38.66	.88
Artistic T1	36.37	34.54	.82
Artistic T2	27.67	33.98	.86
Artistic T3	25.41	32.54	.85
Social T1	75.57	29.94	.79
Social T2	65.83	34.01	.81
Social T3	63.33	35.52	.83
Enterprising T1	41.20	29.93	.70
Enterprising T2	42.43	33.10	.78
Enterprising T3	45.84	33.94	.79
Conventional T1	20.11	23.39	.62
Conventional T2	19.92	25.77	.72
Conventional T3	23.85	27.09	.71

Note. $N_{T1} = 833$; $N_{T2} = 718$; $N_{T3} = 545$.

Study Environment Interest Profiles

For the RIASEC profiles of grade 9 study environments, we adopted the established profiles developed by Dutry et al. (*In Preparation*). Dutry et al. used an incumbent approach to create an interest profile for each cell in the Flemish matrix of study programs. This matrix classifies individual study programs by study domain[33] and finality[34] based on content similarities. Additionally, there are twelve cross-domain programs within the academic

[33]There are eight study domains: (1) Language and Culture, (2) STEM, (3) Art and Creation, (4) Agriculture and Horticulture, (5) Economy and Organization, (6) Society and Well-being, (7) Sport, and (8) Food and Catering.
[34]There are three finalities: the academic finality prepares for higher education, the vocational finality prepares for workforce entry, and the dual finality prepares for both.

finality, five of which are classified as niche. In total, Dutry et al. established 28 profiles[35], reflecting grouped study programs.

Procedure

Participants completed the SIMON-jr interest questionnaire as part of the longitudinal Talentcenter Study. The survey took place in participants' classrooms during regular school hours, using internet-enabled devices such as laptops, computers, or tablets. After a brief explanation, participants completed the questionnaire independently in Qualtrics (https://www.qualtrics.com/). A class teacher and a supervisor from the research university were present to oversee the process. Participants faced no time limits. We repeated the same procedure the following waves. For students who had changed schools, we sent the questionnaire individually via mail with a unique link to ensure accurate data matching. In the final measurement wave, school principals organized the survey due to increased dispersion of students across different classes. Most schools conducted the questionnaire in classrooms with teacher supervision, while some allowed students to complete it at home. For students who had changed schools, we followed the same procedure as in the second wave.

Although participants with incomplete data did not contribute to the analysis of within-person variation in longitudinal multilevel models, they still informed the estimation of fixed effects (Pässler & Hell, 2020; Singer & Willett, 2003).

Analyses

Analyses were conducted in SPSS (version 29.0; IMB Corp., 2022) and R (version 4.3.3; R Core Team, 2024). All R analysis codes as well as the variance-covariance matrix is available at: https://osf.io/2fvy8/?view_only=78eb66cabd544088aba1b7528b072cf3.

[35]Two cells within the domain x finality matrix did not contain any study program, and therefore had no profile. One study program was classified into two cells, each of which contained no other programs; consequently, a single profile was used for this program instead of separate profiles for each cell. For one specific cell (Sport in vocational finality), which includes only niche programs, Dutry et al. (*In Preparation*) did not establish a profile. Similarly, no profiles were established for the five niche cross-domain study programs within the academic finality.

We assessed rank-order stability using Spearman's rho correlations and measured the stability of individual RIASEC profiles with Pearson's correlations. Profile differentiation was operationalized as the standard deviation of an individual's RIASEC scores relative to their personal RIASEC mean. Specifically, we calculated each individual's mean score across the six scales, determined the variance of each scale score from this mean, averaged these variances, and took the square root of the average. This approach is considered optimal for evaluating profile differentiation, as it accounts for the entire interest profile and is independent of profile elevation (Fouad and Mohler, 2004).

The PE interest fit between students' interests and their chosen study program in grade 9 was measured using profile correlation, which compares students' individual interest profiles (at T1, T2, and T3) with the interest profile of their chosen study program at T3. For the program profiles, we utilized the RIASEC profiles of grade 9 study environments developed by Dutry et al. (*In Preparation*), using the incumbent method. The PE interest fit with unchosen study programs in grade 9 was calculated by averaging the correlation between students' interest profiles and all RIASEC profiles of grade 9 study environments, excluding their own program's profile.

To examine the effects of time and gender on the development of RIASEC dimensions, profile differentiation, and PE interest fit, we used multilevel mixed models. The variable *Measurement wave* (categorical) served as a Level 1 predictor, while *Gender* (categorical) was included as a Level 2 variable. Following the approach of Pässler and Hell (2020), we first applied an unconditional means model to calculate the Intra-Class Correlation (ICC) coefficient, assessing the suitability of a multilevel modeling approach. We then employed conditional random-intercept models to investigate the effect of time, as well as potential gender differences in students' initial status (intercepts) and developmental trajectories (slopes). To further investigate specific patterns of change over time, we conducted Bonferroni-corrected pairwise contrasts on significant interactions.

We conducted all multilevel analyses using the lmerTest package in R (version 3.1.3; Kuznetsova et al., 2017). We used restricted maximum likelihood estimation (REML) for parameter estimates (Dedrick et al., 2009); pairwise contrasts were tested with the emmeans package (version 1.10.6; Lenth et al., 2024), and visualizations were generated using the ggplot2 package (version 3.1.5; Wickham, 2016).

Results

Vocational Interest Development

The results of the analyses on mean-level changes in vocational interests are provided in Appendix 5B due to space constraints.

Rank-order and Profile Stability

Table 2 presents the Spearman's rho correlations measuring rank-order stability of the RIASEC interest dimensions, based on participants who participated in all three waves. The highest rank-order stability was observed between T2 and T3 (mean $r = .69$), while the lowest rank-order stability occurred between T1 and T3 (mean $r = .52$). The mean correlation between T1 and T2 was $r = .56$.

Additionally, the table displays rank-order stability separately for boys and girls. Notable gender differences in stability are observed in the Realistic and Artistic dimensions: boys demonstrate higher stability in the Realistic dimension, whereas girls exhibit greater stability in the Artistic dimension. Furthermore, for the Conventional and Enterprising dimensions, boys show lower stability than girls from T1 to T3 and from T1 to T2. This gender difference, however, appears to diminish from T2 to T3.

Table 2

Rank-Order Stability of the RIASEC Interest Dimensions

	Time interval	Spearman's rho Overall	$\Delta r_{overall}$ = r(T2,T3) – r(T1,T2)	Spearman's rho Boys	Spearman's rho Girls
Realistic	T1 to T2	.65**		.57**	.46**
	T2 to T3	.74**	.09	.71**	.58**
	T1 to T3	.58**		.49**	.36**
Investigative	T1 to T2	.60**		.59**	.62**
	T2 to T3	.76**	.16	.76**	.75**
	T1 to T3	.59**		.57**	.63**
Artistic	T1 to T2	.59**		.47**	.49**
	T2 to T3	.71**	.12	.59**	.71**
	T1 to T3	.51**		.33**	.40**
Social	T1 to T2	.56**		.45**	.46**
	T2 to T3	.66**	.10	.58**	.54**
	T1 to T3	.52**		.38**	.40**
Enterprising	T1 to T2	.48**		.38**	.49**
	T2 to T3	.67**	.19	.63**	.66**
	T1 to T3	.47**		.38**	.46**
Conventional	T1 to T2	.46**		.41**	.49**
	T2 to T3	.58**	.12	.57**	.60**
	T1 to T3	.42**		.33**	.52**

Note. N_{overall} = 478; N_{boys} = 248; N_{girls} = 228.

To assess the stability of an individual's RIASEC profile over time, Pearson's correlations were computed. The stability of the profile increased across the measurement waves, with mean correlations of .67 (*SD* = .29) between T1 and T2, of .76 (*SD* = .25) between T2 and T3, and of .61 (*SD* = .33) between T1 and T3. Notably, the highest profile stability was observed between T2 and T3.

When examining the data separately for boys and girls (Table 3), a similar pattern emerged: the highest stability was observed between T2 and T3, and the lowest between T1 and T3. It is noteworthy that girls demonstrated higher stability across all intervals compared to boys.

Table 3

RIASEC Profile Stability by Gender

RIASEC profile correlation	*M (SD)* Boys	*M (SD)* Girls
T1 to T3	.54 (.34)	.68 (.31)
T1 to T2	.62 (.30)	.74 (.28)
T2 to T3	.74 (.26)	.79 (.23)

Differentiation of Interest Profiles

Table 4 presents the mean profile differentiation per measurement wave, including both overall values and values reported separately for boys and girls. To assess the effect of time, we first calculated the ICC, which indicated that 41% of the variance in profile differentiation is attributable to differences between students. This supports the use of a multilevel model to analyze the development of profile differentiation.

The conditional random-intercept model (Table 5), which included time, gender and the interaction between time and gender, revealed a significant gender difference, with girls showing more profile differentiation than boys (B = 4.19, $p < .001$). However, no significant effect of time was observed, nor was there evidence of an interaction between time and gender.

Table 4

Mean Profile Differentiation across Three Waves

Profile differentiation	*M (SD)* Overall	*M (SD)* Boys	*M (SD)* Girls
T1	31.73 (7.63)	29.27 (7.45)	34.39 (6.93)
T2	31.72 (8.23)	30.19 (8.64)	33.42 (7.46)
T3	31.63 (8.56)	29.75 (8.98)	33.71 (7.60)

Note. $N_{overall} = 478$; $N_{boys} = 248$; $N_{girls} = 228$.

Table 5

Results of the Multilevel Model for Profile Differentiation

	β	SE	*t*	*p*
Fixed Effects				
Intercept	29.25	.38	76.75	<.001
Wave1	.20	.46	.43	.67
Wave2	-.04	.51	-.07	.94
Gender1	4.19	.56	7.43	<.001
Wave1:Gender1	-.86	.68	-1.27	.20
Wave2:Gender1	-.30	.74	-.41	.68
Random Effects	Variance	SD		
ID (Intercept)	24.57	4.96		
Residual	41.53	6.44		
Marginal R^2	.05			
Conditional R^2	.40			

Note. The fixed effects intercept represents the baseline value of the level of profile differentiation for boys at T1.

Development of PE Interest Fit

Table 6 displays the mean correlation fit over time with the chosen study program at T3, as well as the mean correlation fit with the unchosen study programs at T3. These values are reported both overall and separately for boys and girls. As in the previous analysis, we first calculated the ICCs to assess the effect of time. The results showed that 51% of the variance in both PE interest fit with the chosen study program and PE interest fit with unchosen study programs were attributable to differences between students. These findings again support the use of a multilevel model to examine the development of PE interest fit.

The conditional random-intercept models (Table 7), which included time, gender and the interaction between time and gender, revealed a significant main effect of gender on PE interest fit with both the chosen and unchosen study programs. Time (wave) had a significant main effect on PE fit with unchosen study programs, but no effect on PE fit with the chosen study program at T3. Additionally, significant interaction effects between time and gender were found in both models, indicating that the rate of change over time differed between genders. To explore these patterns of change further, pairwise contrasts were conducted (Tables 8 and 9).

Hypothesis Testing

H1 posited that PE interest fit with the chosen study program would improve progressively over time. The data did not support this hypothesis, as time did not significantly affect PE fit with the chosen study program for either boys or girls, leading to the rejection of H1. H2 proposed that PE fit with unchosen study programs would diminish over time. The results partially supported this hypothesis, as a significant negative effect of time on PE fit with unchosen study programs was observed, but only for boys. The decline for girls was not statistically significant. Lastly, H3 predicted that girls would exhibit higher PE fit with the chosen study program compared to boys, which was supported by the data. However, this difference was not significant at T1. Figures 1 and 2 depict the PE fit trajectories over time, separated by gender.

Table 6

PE Fit with Chosen and Unchosen Study Programs

		M (SD) Overall	*M (SD)* Boys	*M (SD)* Girls
PE fit with chosen program	T1	.58 (.31)	.55 (.33)	.62 (.29)
	T2	.60 (.31)	.55 (.32)	.65 (.28)
	T3	.59 (.32)	.52 (.35)	.66 (.28)
PE fit with unchosen programs	T1	.39 (.25)	.31 (.28)	.48 (.17)
	T2	.35 (.27)	.25 (.28)	.46 (.21)
	T3	.31 (.31)	.19 (.32)	.45 (.23)

Note. $N_{overall} = 476$; $N_{boys} = 247$; $N_{girls} = 227$.

Table 7

Results of the Multilevel Models for PE Interest Fit

	PE interest fit with chosen study program				*PE interest fit with unchosen study programs*			
	β	SE	*t*	*p*	β	SE	*t*	*p*
Fixed Effects								
Intercept	.54	.02	30.26	<.001	.32	.01	22.16	<.001
Wave1	-.00	.02	-.01	.99	-.08	.02	-4.84	<.001
Wave2	-.03	.02	-1.53	.13	-.12	.02	-7.51	<.001
Gender1	.07	.03	2.51	.01	.16	.02	7.49	<.001
Wave1:Gender1	.04	.03	1.40	.16	.06	.02	2.57	.01
Wave2:Gender1	.07	.03	2.39	.02	.08	.02	3.51	<.001
Random Effects	Var.	SD			Var.	SD		
ID (Intercept)	.05	.22			.03	.17		
Residual	.05	.23			.04	.19		
Marginal R^2	.03				.15			
Conditional R^2	.51				.54			

Table 8

Pairwise Contrasts for Fit with Chosen Study Program across Waves and Gender

Contrast		Estimate	SE	*p*	*cohen's d*
T1	boys - girls	-.07	.03	.182	-.30
T2	boys - girls	-.10	.03	.001	-.46
T3	boys - girls	-.13	.03	<.001	-.59
Boys	T1 - T2	.00	.02	1.00	.00
	T2 -T3	.03	.02	1.00	.13
	T1 - T3	.03	.02	1.00	.13
Girls	T1 - T2	-.04	.02	.86	-.17
	T2 -T3	.00	.02	1.00	.00
	T1 - T3	-.04	.02	1.00	-.16

Note. *Cohen's d* reflects the effect size after controlling for individual-level variations and does not account for multiple testing.

Table 9

Pairwise Contrasts for Fit with Unchosen Study Programs across Waves and Gender

Contrast		Estimate	SE	*p*	*cohen's d*
T1	boys - girls	-.16	.02	<.001	-.84
T2	boys - girls	-.22	.02	<.001	-1.15

T3	boys - girls	-.24	.02	<.001	-1.27
Boys	T1 - T2	.08	.02	<.001	.40
	T2 -T3	.04	.02	.09	.23
	T1 - T3	.12	.02	<.001	.63
Girls	T1 - T2	.02	.02	1.00	.09
	T2 -T3	.02	.02	1.00	.11
	T1 - T3	.04	.02	.40	.20

Note. *Cohen's d* reflects the effect size after controlling for individual-level variations and does not account for multiple testing.

Figure 1

PE Interest Fit with Chosen Study program over Time per Gender

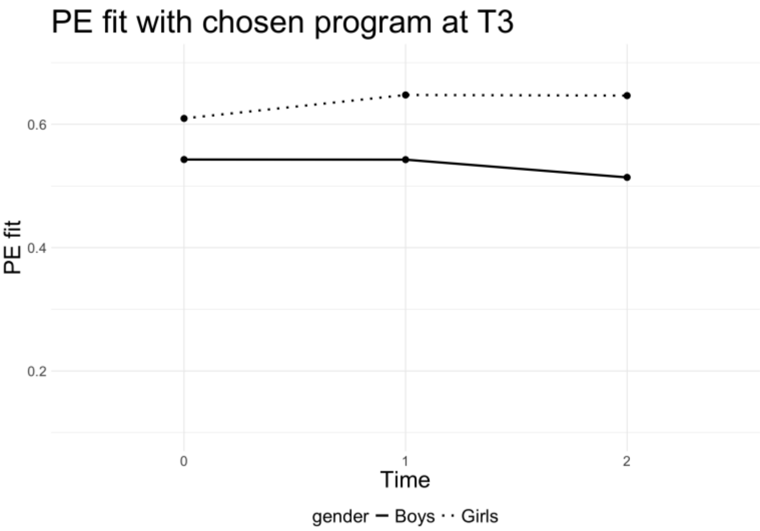
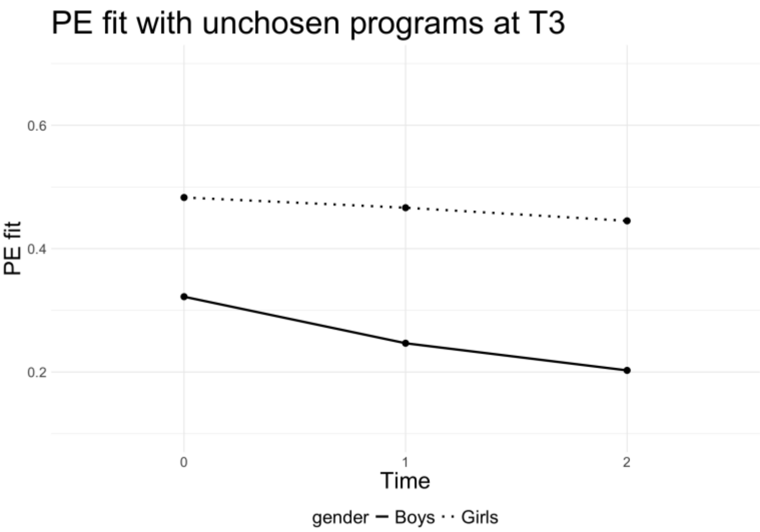


Figure 2

PE Interest Fit with Unchosen Study Programs over Time per Gender



Given that our results suggest an increasing difference between the fit with the chosen program and the fit with the unchosen programs over time, we conducted a post-hoc analysis to examine this difference further. Specifically, we created three new variables, each representing the difference between the fit with the chosen program and the fit with the unchosen programs for each wave. Consistent with previous analyses, we first calculated the ICC to assess the effect of time. The analysis revealed that 58% of the variance in the difference between the fit with the chosen and unchosen programs was attributable to differences between students. This finding underscores the appropriateness of using a multilevel model to investigate the development of this difference.

The conditional random-intercept model (Table 10), which included time, gender, and the interaction between time and gender, revealed significant main effects of both gender and time on the difference between the fit with the chosen and unchosen study programs. No significant interaction effect between time and gender was observed, indicating that the rate of widening of the gap was consistent across genders.

Table 10

Results of the Multilevel Model for the Difference between Fit with the Chosen and Fit with the Unchosen Programs

	β	SE	*t*	*p*
Fixed Effects				
Intercept	.22	.02	13.31	<.001
Wave1	.07	.02	4.79	<.001
Wave2	.09	.02	5.67	<.001
Gender1	-.09	.02	-3.86	<.001
Wave1:Gender1	-.02	.02	-.83	.41
Wave2:Gender1	-.01	.02	-.64	.52
Random Effects	Variance	SD		
ID (Intercept)	.05	.23		
Residual	.04	.19		
Marginal R^2	.04			
Conditional R^2	.60			

Note. The fixed effects intercept represents the baseline value of the difference between the fit with the chosen and the fit with the unchosen programs for boys at T1.

To estimate the effect size of the widening difference between the fit with the chosen program and the fit with the unchosen programs, we conducted pairwise contrasts for time (Table 11), given that the interaction between time and gender was not significant. The results indicate that this difference increased over time for both boys and girls, with a small effect size.

Table 11

Pairwise Contrasts for Difference between Fit with Chosen and Unchosen Programs across Waves

Contrast	Estimate	SE	*p*	*cohen's d*
T1 - T2	-.07	.01	<.001	-.34
T2 -T3	-.02	.01	.38	-.09
T1 - T3	-.08	.01	<.001	-.44

Note. *Cohen's d* reflects the effect size after controlling for individual-level variations and does not account for multiple testing.

Discussion

The present study aimed to examine the longitudinal development of PE interest fit during early adolescence within the Flemish educational system. To frame our findings within the existing literature and contribute to the understanding of the dynamics of PE interest fit, we also analyzed patterns of change and stability in vocational interests over time. Although these analyses were not part of the primary research questions, they provide critical context and strengthen the interpretation of our results. Our findings reveal that PE fit with the chosen program remains stable over time, which is notable given the changes observed in singular interest dimensions. In contrast, PE fit with unchosen programs decreases over time, although only at a significant level for boys. For both boys and girls, the difference between the fit with chosen and unchosen programs becomes increasingly pronounced, primarily driven by the decline in fit with unchosen options. These findings highlight gender differences in the development of PE interest fit and offer valuable insights for advancing vocational interest theories across diverse educational contexts. Below we provide a more detailed overview and interpretation of the results.

Vocational Interest Development

In terms of rank-order stability, our findings revealed moderate to high correlations. Compared to the results of Gfrörer et al. (2022), our study observed slightly higher stability coefficients. However, it is important to note that our sample included students up to grade 9 (ages 14–15), whereas Gfrörer et al. (2022) only followed students through grade 8 (ages 13–14). When focusing on overlapping age groups, the stability levels in our findings are highly comparable. Consistent with the expectation that vocational interests become increasingly stable during adolescence, the inclusion of older students in our sample likely contributed to the slightly higher stability observed.

In terms of profile stability, we observed relatively high correlations across time and within participants, aligning with prior research findings (e.g., population estimate $\rho = .70$ in Low et al., 2005; *r*'s ranging from .68 to .81 in Xu & Tracey, 2016), suggesting that individuals generally maintained consistent patterns of interest, with relatively few experiencing notable shifts in their interest configurations. Notably, girls exhibited greater profile stability across all intervals compared to boys, a pattern consistent with earlier studies (Tracey et al., 2005; Xu & Tracey, 2016).

Concerning patterns of change, we observed a general decline in the Realistic, Investigative, Artistic and Social interest dimensions, alongside a slight increase in the Enterprising dimension. The magnitude of these changes varied by gender, time interval, and specific RIASEC dimension. These findings are largely in line with previous literature (Gfrörer et al., 2022; Hoff et al., 2018).

Consistent with previous research, boys displayed stronger Realistic and Investigative interests than girls. However, while the gender difference in Investigative interests was significant at T1, the difference disappeared in later waves, contrary to general trends reported in literature (Gfrörer et al., 2022; Su et al., 2009). As expected, girls demonstrated stronger Artistic and Social interests, with the gender gap in Social interests widening over time. In contrast to the findings of Hoff et al. (2018) and Gfrörer et al. (2022), we found no increasing gender difference in the Realistic dimension.

PE Interest Fit Development

Building on the foundational work of Holland (1997) and Schneider (1987), which suggest that individuals gravitate toward environments fitting with their evolving interests, we hypothesized that PE interest fit would progressively improve throughout adolescence. However, our findings did not support this hypothesis, as we observed no significant effect of time on PE interest fit with the chosen study program in grade 9 for either boys or girls. Contrary to expectations, adolescents did not exhibit a clearer fit between their evolving interests and their chosen study program as they progressed through adolescence. This result could suggest that factors beyond selection and socialization (Hanna et al., 2021)—such as social influences, academic pressures, or external expectations—may play a more prominent role in shaping students' program choices. Alternatively, the lack of improvement in PE interest fit could indicate that the development of a coherent self-concept and a differentiated RIASEC profile is a more gradual or complex process than previously anticipated (Slot et al., 2021).

The finding that PE interest fit with the chosen study program in secondary education (mean PE fit = .59) is lower than the PE fit typically observed in higher education (mean PE fit = .70; Schelfhout et al., 2019) supports both of the above interpretations. Students in secondary education are likely more influenced by parental guidance (Goosen et al., 2017), leading to choices that may not fully align with their personal interests. In addition, research suggests that vocational interests continue to evolve through late adolescence (Hoff et al., 2018). This ongoing developmental process might make it challenging for students in secondary education to select programs that fully align with their interests. An alternative explanation for the lower PE fit in secondary education could be the nature of the educational environments in Flanders. While secondary education programs are organized around ability and interest domains, many remain quite general, preserving access to virtually all domains in higher education. As a result, the environmental RIASEC profiles are less differentiated compared to the more specialized environments in higher education. This generality implies unavoidable moderation of PE interest fit. Future research could explore this further by examining PE interest fit in more specialized educational settings, such as vocational education programs, which are often designed to prepare students for specific professions. As a preliminary exploration, we compared PE fit at T3 between students in academic and vocational tracks. Descriptive analysis revealed that students in the vocational track achieved a higher PE fit ($M = .62$; $SD = .28$) compared to students in the academic track ($M = .57$; $SD = .32$), with the academic track showing similar PE fit levels to students in the dual track (preparing for both workforce entry and higher education; $M = .57$; $SD = .36$). However, it is important to note the substantial disparity in sample sizes between the groups: 313 students in the academic track, 102 in the vocational track, and 124 in the dual track. Therefore, we recommend further research to more comprehensively examine the relationship between program type and (development of) PE interest fit.

Despite the absence of improvement in PE fit throughout adolescence and its lower level compared to higher education, the observed PE fit in secondary education is still relatively high. Many students appear to select programs that fit reasonably well with their interests, underscoring the importance of supporting interest-driven educational choices, even within general education contexts. Furthermore, while a general decline is observed in most vocational interest dimensions, with varying magnitudes of change across the RIASEC dimensions, it is noteworthy that interest fit does not deteriorate. This suggests that the study program may be well-aligned with students' core vocational interests, ensuring that, even as overall interest levels decrease, the program continues to match their most prominent interests. This finding underscores the robustness of PE interest fit as a valuable tool for supporting young adolescents in their study choices, even amidst developmental changes.

We hypothesized that the PE interest fit with unchosen study programs would diminish over time, and our findings partially supported this hypothesis. Specifically, we observed a significant negative effect of time on the PE interest fit with unchosen study programs among boys. However, while a similar trend was observed for girls, the effect was not statistically significant. This decline in fit with unchosen programs aligns with Gottfredson's theory of circumscription and compromise (1981), which suggests that as individuals commit to specific career paths and refine their self-concept, they increasingly distance themselves from alternative options they initially considered but ultimately did not choose. Although the decline in fit with unchosen programs was not significant for girls, the post-hoc analysis revealed that girls—just like boys—showed a more pronounced difference over time between the fit with their chosen and unchosen programs. This finding suggests that Gottfredson's theory may also apply to girls, albeit with a less pronounced effect.

One possible explanation for this gender difference may lie in the fact that girls generally exhibit stronger PE fit. It stands to reason that a strong fit with one program could correlate with a good fit for others, particularly when the programs are content-related. In our study, girls demonstrated higher PE fit with their chosen study programs compared to boys, which could account for the less pronounced decline in fit with unchosen programs. If girls fit better with their chosen programs, this may also lead to relatively higher fit with unchosen programs, thus mitigating the observed decline. This relationship between fit with chosen and unchosen programs may reduce the need to search for alternative explanations for the gender difference in the decline of fit. Therefore, the critical question may not be why girls experience a smaller decline in fit with unchosen programs, but rather why girls exhibit stronger overall fit to begin with.

A potential factor contributing to girls' stronger PE fit could be that girls engage more actively in career exploration than boys (e.g., Demulder et al., 2022; Driesel-Lange, 2011; Lazarides et al., 2016). Indeed, career exploration plays a crucial role in developing career identities, aligning one's self-concept with the environment, and supporting vocational development (Jiang et al., 2019; Savickas, 2002, 2013). Lazarides et al. (2016) suggest that the greater career exploration among girls may stem from a need to reflect on gender-atypical career paths. In many countries, careers stereotypically considered “female” are often associated with lower income, and the labor market continues to be characterized by sex-segregated occupations (Federal Agency for Statistics, 2010). Additionally, research has shown that female students tend to have clearer knowledge of their interests and potential study choices, and they make more effective use of that knowledge compared to their male counterparts (Germeijs & Verschueren, 2006; Schelfhout et al., 2021b). Given these findings, it is essential to provide both boys and girls with adequate support in their career-related self-exploration, helping them navigate their interests and make informed choices that fit with their aspirations.

In conclusion, our findings reveal that PE fit with the chosen program remains stable over time, which is notable given the changes observed in singular interest dimensions. In contrast, PE fit with unchosen programs decreases over time, particularly among boys. For both boys and girls, the difference between the fit with chosen and unchosen programs becomes increasingly pronounced, primarily driven by the decline in fit with unchosen options. These findings highlight the complex and multifaceted nature of how adolescents develop their academic identities, suggesting that both individual processes and structural factors play significant roles in shaping these trajectories.

Practical Implications

Our findings underscore the importance of incorporating vocational interest assessment tools into the career decision-making processes of early adolescents. Despite developmental changes in vocational interests across most RIASEC dimensions, the stability of PE fit with the environment chosen in grade 9 suggests that these program choices already fit with students' core interests as early as grade 7, indicating that vocational interests can be used for study orientation. This finding is particularly relevant for educational systems like the Flemish model, where students must make early decisions—such as choosing a school or elective course—that shape their future academic and vocational trajectories. By leveraging vocational interest assessments, educators and policymakers can provide guidance to students and their families, helping them navigate these choices with greater confidence. By fitting students with environments that reflect their intrinsic interests, these tools can promote engagement and long-term success (de Vries et al., 2023). However, this study did not investigate how PE fit relates to positive outcomes such as academic achievement, satisfaction, and program retention. Investigating these potential relations in future research is essential to understanding the broader implications of PE interest fit (or its absence) during adolescence.

At the same time, our findings reveal that vocational interests and RIASEC profile configurations are not fixed. As demonstrated in previous research (Pässler & Hell, 2020) and confirmed by our findings, gender differences in vocational interests are evident from an early age and persist over time, contributing to gender disparities in STEM careers (Eccles & Wang, 2015; Kossek et al., 2017). Furthermore, our study revealed that boys showed greater stability in the Realistic dimension, while girls exhibited more stability in the Artistic dimension. These differences may arise from unequal exposure to relevant activities, as gendered experiences can shape interest development (Gfrörer et al., 2021). For instance, when girls are less exposed to Realistic activities, they may develop weaker interest in these fields (Su et al., 2019). These findings suggest that targeted educational initiatives could foster girls' interest in the Realistic dimension. By introducing such initiatives during late childhood and early adolescence—crucial years for shaping career interests and gender roles (Gottfredson, 1981)—educators and policymakers have an opportunity to foster more balanced career aspirations, ultimately encouraging greater female participation in STEM fields.

Strengths, Limitations and Suggestions for Future Research

A key strength of this study is that it provides novel insights into how PE fit evolves in secondary education—an area that has received limited attention. By examining fit relative to a single environment chosen at T3, we were able to capture how individuals grow into their selected context over time, offering a unique perspective on the development of PE fit. However, this approach differs from designs that assess fit dynamically across multiple environments at each measurement point. Although the educational system in Flanders constrained us to this approach—particularly given our focus on early adolescence—this may have underestimated the role of socialization processes during this developmental stage. Future studies should consider longitudinal designs that incorporate the influence of specific environments at each time point to further refine our understanding of PE fit dynamics.

Second, this study provided a broad, descriptive account of how PE interest fit evolves over time, offering a foundational perspective on its development in early adolescence. However, we did not investigate the underlying factors driving the development of PE interest fit. It is possible that shifts in PE fit at the individual level balanced each other out, leading to no significant differences in overall fit. For example, some students may experience positive changes in fit due to factors like increased interest in their chosen field or positive experiences with elective courses, while others may experience declines in fit due to external pressures. As a result, we may have missed valuable insights into the processes that shape the development of PE fit for different individuals. Future research should focus on identifying and examining the specific factors that shape the development of PE interest fit and exploring how these factors interact to influence individual trajectories. Additionally, understanding how these trajectories relate to positive outcomes, such as academic success, satisfaction, and retention in the program, could provide valuable insights into the broader implications of PE interest fit during adolescence.

Conclusion

In conclusion, this study contributes to the understanding of the longitudinal development of PE interest fit during early adolescence. The results demonstrated that students already exhibit a substantial fit with their grade 9 program choice at the start of secondary education. While students' interests evolve during adolescence, this evolution does not necessarily result in stronger fit with their chosen study program, as the fit remains stable over time. Instead, students—particularly boys—tend to progressively lose interest in the options they have excluded, reflecting a narrowing of focus. Moreover, for both boys and girls, the difference between the fit with chosen and unchosen programs becomes increasingly pronounced over time, primarily driven by the decline in fit with unchosen options. These findings underscore the complex interplay between individual processes and structural factors in shaping adolescents' academic identities. Future research should explore the underlying processes and predictors of PE fit development in early adolescence.

References

Allen J., & Robbins S. (2010). Effects of interest–major congruence, motivation, and academic performance on timely degree attainment. *Journal of Counseling Psychology, 57,* 23–35. https://doi.org/10.1037/a0017267 PMID: 21133558

Bleidorn, W., Schwaba, T., Zheng, A. Q., Hopwood, C. J., Sosa, S. S., Roberts, B. W., & Briley, D. A. (2022). Personality Stability and Change: A Meta-Analysis of Longitudinal Studies. *Psychological Bulletin, 148,* 7-8. https://doi.org/ 10.1037/bul0000365

Demulder, L., Verschueren, K., Lacante, M., & Donche, V. (2022). Identifying exploration profiles for higher education and their relationship with different student variables and outcomes. *European Journal of Psychology of Education.* https://doi.org/10.1007/s10212-022-00641-w

de Vries, N., Meeter, M., & Huizinga, M. (2024). Does interest fit between student and study program lead to better outcomes? A meta-analysis of vocational interest congruence as predictor for academic success. *Educational Research Review, 44,* 100619. https://doi.org/10.1016/j.edurev.2024.100619

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., ... Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research, 79,* 69–102. https://doi.org/10.3102/0034654308325581

Driesel-Lange, K. (2011). *Berufswahlprozesse von Mädchen und Jungen: Intervention-smöglichkeiten zur Förderung geschlechtsunabhängiger Berufswahl* [Girls' and boys'vocational choices: Interventions to facilitate gender-independent vocational choices]. Berlin: Lit

Dutry, M., Schelfhout, S., Derous, E., Duyck, W., & Dirix, N. (n.d.). Optimizing career guidance in secondary education: Exploring person-environment interest fit with SIMON-junior interest inventory. *Manuscript in Preparation.*

Eccles, J. S., & Wang, M.-T. (2015). What motivates females and males to pursue careers in mathematics and science? *International Journal of Behavioral Development, 40*, 100–106. https://doi.org/10.1177/0165025415616201.

Einarsdóttir, S., & Rounds, J. (2020). A quantitative review of gender differences in vocational interests in Iceland: Pervasive and persistent. *Nordic Journal of Transitions, Careers and Guidance, 1,* 10–26. https://doi.org/10.16993/njtcg.29

Ertl, B., Hartmann, F. G., & Wunderlich, A. (2023). Stability of vocational interests during university studies. *Journal of Individual Differences, 44, 143*-162. https://doi.org/10.1027/1614-0001/a000392

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Review System for Evaluating Test Quality.* Boom Uitgevers Amsterdam. Retrieved from https://psynip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf

Federal Agency for Statistics. (2010). *Männer und Frauen in verschiedenen Lebensphasen* [Men and women in different developmental stages]. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/Haushalte-Familien/

Fouad, N. A., & Mohler, C. J. (2004). Cultural validity of Holland's theory and the Strong Interest Inventory for five racial/ethnic groups. *Journal of Career Assessment, 12,* 423– 439. http://dx.doi.org/10.1177/10690 72704267736

Germeijs, V., & Verschueren, K. (2006). High school students' career decision-making process: Development and validation of the study choice task inventory. *Journal of Career Assessment, 14,* 449–471. https://doi.org/10.1177/1069072706286510

Gfrörer, T., Stoll, G., Rieger, S., Trautwein, U., & Nagengast, B. (2022). The development of vocational interests in early adolescence: Stability, change, and state-trait components. European Journal of Personality, 36, 942. https://doi.org/10.1177/08902070211035630

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology, 28,* 545–579. https://doi.org/10.1037/0022-0167.28.6.545

Granillo-Velasquez, K. E., Hoff, K. A., Hanna, A., Oswald, F. L., & Morris, M. L. (2024). Comparing methods of measuring interest fit: A large prediction study with career choice satisfaction. *International Journal of Selection and Assessment,* 1–13. https://doi.org/10.1111/ijsa.12506

Hanna, A., Briley, D., Einarsdóttir, S., Hoff, K., Rounds, J. (2021). Fit gets better: A longitudinal study of changes in interest fit in educational and work environments. *European Journal of Personality, 35,* 557-580. https://doi.org/ 10.1177/08902070211014022

Hoff, K. A., Briley, D. A., Wee, C. J. M., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychological Bulletin, 144,* 426–451. https://doi.org/ 10.1037/bul0000140

Hoff, K. A., Song, Q. C., Wee, C. J., Phan, W. M. J., & Rounds, J. (2020). Interest fit and job satisfaction: A systematic review and meta-analysis. *Journal of Vocational Behavior, 123,* Article 103503. https://doi.org/10.1016/j.jvb.2020.103503

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL, US: Psychological Assessment Resources.

Jiang, Z., Newman, A., Le, H., Presbitero, A., & Zheng, C. N. (2019). Career exploration: A review and future research agenda. *Journal of Vocational Behavior, 110,* 338-356. https://doi.org/ 10.1016/j.jvb.2018.08.008

Keijsers, L., & Poulin, F. (2013). Developmental changes in parent–child communication throughout adolescence. *Developmental Psychology, 49,* 2301–2308. https:// doi.org/10.1037/a0032217

Kossek, E. E., Su, R., & Wu, L. (2017). “Opting out” or “pushed out”? Integrating perspectives on women's career equality for gender inclusion and interventions. *Journal of Management, 43,* 228–254. https://doi.org/10.1177/0149206316671582.

Krapp, A. (2002). Structural and dynamic aspects of interest development: Theoretical considerations from an ontogenetic perspective. *Learning and Instruction, 12,* 383–409. https://doi.org/10.1016/S0959-4752(01) 00011-1

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of statistical software, 82,* 1-26. https://doi.org/10.18637/jss.v082.i13

Lasselle, L., Schelfhout, S., Fonteyne, L., Kirby, G., Smith, I., & Duyck, W. (2021). An examination of gender imbalance in Scottish adolescents’ vocational interests. PLOS ONE, 16. https://doi.org/10.1371/journal.pone.0257723

Lazarides, R., Rohowski, S., Ohlemann, S., & Ittel, A. (2016). The role of classroom characteristics for students' motivation and career exploration. *Educational Psychology, 36,* 992-1008. https://doi.org/10.1080/01443410.2015.1093608

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*,79-122. https://doi.org/10.1006/jvbe.1994.1027

Lenth, R. V., Banfai, B., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., … Singmann, H. (2024). *Estimated Marginal Means, aka Least-Squares Means* [R package]. 10.32614/CRAN.package.emmeans

Low, K. S. D., Yoon, M., Roberts, B. W., & Rounds, J. (2005). The stability of vocational interests from early adolescence to middle adulthood: A quantitative review of longitudinal studies. Psychological Bulletin, 131, 713–737. https://doi.org/10.1037/0033-2909.131. 5.713

Nauta, M. M. (2010). The development, evolution, and status of Holland's theory of vocational personalities: reflections and future directions for counseling psychology. *Journal of Counseling Psychology, 57,* 11–22. https://doi.org/10.1037/a0018213

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2012). Vocational interests and performance: A quantitative summary of over 60 years of research. *Perspectives on Psychological Science, 7,* 384–403. https://doi.org/10.1177/1745691612449021

Morris, M. L. (2016). Vocational interests in the United States: Sex, age, ethnicity, and year effects. *Journal of Counseling Psychology, 63,* 604–615. https://doi.org/10.1037/cou0000164.

Nye, C. D., Su, R., Rounds, J., & Drasgow, F. (2017). Interest congruence and performance: Revisiting recent meta-analytic findings. *Journal of Vocational Behavior, 98,* 138–151. https://doi.org/10.1016/j.jvb.2016.11.002

Nye, C. D., Wille, B., Amory, J., & De Fruyt, F. (2021). Are work activities related to interest change over time? A 22-year longitudinal study. J*ournal of Personality and Social Psychology, 121*, 865–893. https://doi.org/10.1037/pspp0000360

Pässler, K., & Hell, B. (2020). Stability and change in vocational interests from late childhood to early adolescence. *Journal of Vocational Behavior, 121,* 103462. https://doi.org/ 10.1016/j.jvb.2020.103462

R Core Team. (2024). *R: A language and environment for statistical computing* (version 4.3.3) [Software]. Vienna, Austria. Retrieved from http://www.R-project.org/

Roberts, B. W., & Robins, R. W. (2004). Person-environment fit and its implications for personality development: A longitudinal study. *Journal of Personality*, *72*, 89-110. https://doi.org/ 10.1111/j.0022-3506.2004.00257.x

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin, 132,* 1–25. https://doi.org/10. 1037/0033-2909.132.1.1

Rocconi, L. M., Liu, X. Q., & Pike, G. R. (2020). The impact of person-environment fit on grades, perceived gains, and satisfaction: An application of holland's theory. *Higher Education, 80*, 857–874. https://doi.org/10.1007/s10734-020-00519-0

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25,* 54–67. https://doi.org/10.1006/ceps.1999.1020

Salgado, M., Gonzalez, L., & Yanez, A. (2021). Parental Involvement and Life Satisfaction in Early Adolescence. *Frontiers in Psychology, 12,* 628720. https://doi.org/10.3389/fpsyg.2021.628720

Savickas, M. L. (2002). Career construction: A developmental theory of vocational behavior. In D. Brown (Ed.). *Career choice and development* (pp. 255–311). (4th ed.). San Francisco, CA: Jossey-Bass.

Savickas, M. L. (2013). Career construction theory and practice. In S. D. Brown, & R. W. Lent (Eds.). *Career development and counseling: Putting theory and research to work* (pp. 147–183). (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Seghers, M., Boone, S., & Van Avermaet, P. (2019). Classed patterns in the course and outcome of parent-teacher interactions regarding educational decision-making. *Educational Review, 73,* 417-435. https://doi.org/10.1080/00131911.2019.1662771

Slot, E. M., Bronkhorst, L. H., Akkerman, S. F., & Wubbels, T. (2021). Vocational interest profiles in secondary school: Accounting for multiplicity and exploring associations with future-oriented choices. *Journal of Educational Psychology, 113,* 1059-1071. https://doi.org/10.1037/edu0000475

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2019). The effects of vocational interest on study results: Student Person – Environment fit and program interest diversity. *PLoS One, 14*, e0214618, https://doi.org/10.1371/journal.pone.0214618

Schelfhout, S., Bassleer, M., Wille, B., Van Cauwenberghe, S., Dutry, M., Fonteyne, L., … Duyck, W. (2022). Regressed person-environment interest fit: Validating polynomial regression for a specific environment. *Journal of Vocational Behavior, 136.* https://doi.org/10.1016/j.jvb.2022.103748

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2021a). From interest assessment to study orientation: An empirical advice set engine. *The Journal of Experimental Education, 89,* 169–195. https://doi.org/10.1080/00220973.2019.1605327

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., Derous, E., De Fruyt, F., & Duyck, W. (2021b). How interest fit relates to STEM study choice: Female students fit their choices better. *Journal of Vocational Behavior, 129,* 103614. https://doi.org/10.1016/j.jvb.2021.103614

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437-453. doi:10.1111/j.1744-6570.1987.tb00609.x

Soto, C. J., & Tackett, J. L. (2015). Personality traits in childhood and adolescence: Structure, development, and outcomes. *Current Directions in Psychological Science, 24,* 358–362. https://doi.org/10.1177/ 0963721415589345

IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29.0). Armonk, NY: IBM Corp

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford, New York: Oxford University Press

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135*, 859–884. https://doi.org/10.1037/a0017364

Su, R., Stoll, G., & Rounds, J. (2019). The nature of interests: Toward a unifying theory of trait-situation interest dynamics. In C. D. Nye & J. Rounds (Eds.), *Vocational interests: Rethinking their role in understanding workplace behavior and practice* (pp. 11–38). Taylor & Francis/ Routledge. https://doi.org/10.4324/9781315678924-2

Sung, Y. T., Cheng, Y. W., & Hsueh, J. H. (2017). Identifying the career-interest profiles of junior-high-school students through latent profile analysis. *The Journal of Psychology, 151,* 229 –246. http://doi .org/10.1080/00223980.2016.1261076

Tracey, T. J. G. (2002). Development of Interests and Competency Beliefs: A 1-Year Longitudinal Study of Fifth- to Eighth-Grade Students Using the ICA–R and Structural Equation Modeling. *Journal of Counseling Psychology, 49* 148-163. https://doi.org/10.1037//0022-0167.49.2.148

Tracey, T.J.G., Allen, J., & Robbins, S.B. (2012). Moderation of the relation between personen-vironment congruence and academic success: environmental constraint, personal flexibility and method. *Journal of Vocational Behavior, 80,* 38-49. https://doi:10.1016/j.jvb.2011.03.005

Tracey, T. J. G., & Robbins, S. B. (2005). Stability of interests across ethnicity and gender: A longitudinal examination of grades 8 through 12. *Journal of Vocational Behavior, 67,* 335–364. http://doi.org/10.1016/ j.jvb.2004.11.003

Tracey, T. J. G., Robbins, S. B., & Hofsess, C. D. (2005). Stability and change in interests: A longitudinal study of adolescents from grades 8 through 12. *Journal of Vocational Behavior, 66,* 1–25. http://doi.org/ 10.1016/j.jvb.2003.11.002

Tracey, T. J. G., & Sodano, S. M. (2008). Issues of stability and change in interest development. The Career Development Quarterly, 57, 51–62. https://doi.org/10.1002/ j.2161-0045.2008.tb00165.x.

Van Iddekinge, C. H., Putka, D. J., & Campbell, J. P. (2011). Reconsidering vocational interests for personnel selection: The validity of an interest-based selection test in relation to job knowledge, job performance, and continuance intentions. *Journal of Applied Psychology, 96,* 13–33. https://doi.org/10.1037/a0021193.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org

Wille, B., & De Fruyt, F. (2014). Vocations as a source of identity: Reciprocal relations between Big Five personality traits and RIASEC characteristics over 15 years. *Journal of Applied Psychology*, *99*, 262. https://doi.org/10.1037/a0034917

Xu, H., & Li, H. (2020). Operationalize interest congruence: A comparative examination of four approaches. *Journal of Career Assessment, 28,* 571–588. https://doi.org/10.1177/1069072720909825

Xu, H., & Tracey, T. J. G. (2016). Stability and change in interests: A longitudinal examination of grades 7 through college. *Journal of Vocational Behavior, 93,* 129–138. https://doi.org/10.1016/j.jvb.2016.02.002.

Chapter 6.

Discussion

By fostering self-awareness and environmental insight, career exploration contributes to more satisfying career decisions (Hirschi, 2009; Lent & Brown, 2013; Savickas, 2013). Despite the recognized importance of career exploration, many secondary education students face the challenge of making pivotal educational decisions with limited information and guidance (Goosen et al., 2017). This tendency highlights the need for valid, context-specific tools to support students in navigating the complexities of educational and career choices. In Flanders, the establishment of the Flemish Talentcenters aimed to address this need by providing tailored career guidance. Since their inception on April 18, 2023, these centers have served over 9,000 students from more than 200 schools, reflecting the substantial demand for targeted support in career decision-making.

This dissertation documented the development, validation, and application of innovative instruments designed to enhance the study choice process in secondary education, tailored to the needs of the Flemish educational system. Guided by career theories, such as person-environment fit theory (Dawis, 2005; Holland, 1997; Schneider, 1987) and career exploration theory (Jiang et al., 2019), the research emphasized the dynamic interplay between personal and environmental factors in shaping students' career-related behaviors. This final chapter synthesizes the key findings of these studies, situating them within the broader theoretical framework of career exploration. This chapter also highlights the practical implications of the research findings for enhancing educational guidance, particularly through initiatives such as the Flemish Talentcenters, and explores opportunities for future research to further enhance these efforts.

Research Overview

Children's ICAR

Chapter 2 of this dissertation focused on the development and validation of the Children's ICAR (Ch-ICAR), an adaptation of the ICAR project (Revelle et al., 2020) designed to assess cognitive abilities in children. This work addressed a critical gap in the availability of psychometrically robust, freely accessible, and time-efficient cognitive ability measures specifically tailored for younger populations. The chapter outlined the development of the Ch-ICAR, starting with a pilot study conducted with primary school pupils to compile and refine the test items. Following this pilot study, Study 1 evaluated the psychometric properties of the Ch-ICAR in a sample of secondary school pupils. Finally, Study 2 examined the construct validity of the Ch-ICAR by cross-validating it against two established cognitive ability measures: Raven's 2 Progressive Matrices (RPM; McLeod & McCrimmon, 2021) and the Flemish CoVaT-CHC Basic Version (Magez et al., 2015). These cross-validation analyses were conducted using separate samples of primary and secondary school pupils.

The findings strongly support the utility of the full Ch-ICAR as a reliable and valid tool for measuring children's cognitive abilities in research contexts. Study 1 demonstrated the psychometric robustness of the Ch-ICAR through a hierarchical IRT analysis. The remaining 27 items showed adequate discriminatory power with no evidence of differential item functioning (DIF). The model exhibited good fit, consistent with CHC theory, which posits broad cognitive abilities at the lower level and general cognitive ability at the higher level (Flanagan & McDonough, 2018). This refined 27-item subset was used in subsequent analyses and is recommended for future research employing the Ch-ICAR.

Study 1 also demonstrated strong associations between the Ch-ICAR and academic performance, aligning with prior validation studies of the ICAR16 (Condon & Revelle, 2014). Specifically, the Ch-ICAR explained 18% of the within school variance in GPA, 26% of the within school variance in mathematics scores, and 35% of the within school variance in CDR (a standardized basic mathematical skill test) scores, highlighting its relevance as a tool for studying the cognitive foundations of academic achievement.

Study 2 confirmed the construct validity of the Ch-ICAR by demonstrating high correlations with two established intelligence tests. The magnitude of these correlations aligned with findings from prior cross-validation research on the ICAR16 (Young & Keith, 2020). As anticipated, the Ch-ICAR primarily measures Gf (fluid intelligence), making it a reliable tool for assessing non-verbal cognitive ability. Subtest-level analyses, including a correlation matrix and an integrated CFA model, demonstrated the alignment of the subtests with the expected constructs. Number Series and Matrix Reasoning primarily measured Gf; Figural Analogies reflected both Gf and Gv (visuospatial processing); and Verbal Reasoning assessed a combination of Gf and Gc (crystallized intelligence). However, caution is advised when using individual subtests—particularly Matrix Reasoning and Figural Analogies—in isolation, as they showed lower psychometric robustness and weaker connections to broad cognitive abilities. Further replication studies are necessary to confirm the relationships between these subtests and broad cognitive abilities and to refine these components for future applications.

CalibrainVR-MR

Chapter 3 focused on validating CalibrainVR-MR, a virtual reality (VR)-based tool for assessing mental rotation abilities in secondary education students. The study examined the tool's psychometric properties and its alignment with established research linking spatial abilities to STEM performance (Sorby et al., 2018; Wai et al., 2009). By addressing the lack of reliable VR applications for evaluating mental rotation ability, particularly in younger populations, Chapter 3 tackled a significant gap in the field.

The study demonstrated that the final 34-item subset of CalibrainVR-MR achieves good internal consistency. This subset was used for all subsequent analyses, and is recommended for future applications of the tool.

CalibrainVR-MR showed medium correlations with basic mathematical skills, reinforcing the well-documented link between mental rotation abilities and mathematical performance (Xie et al., 2020). Similarly, the observed medium correlations with cognitive ability aligned with prior research, as mental rotation is a recognized component of spatial intelligence (Chen, 2022; Hegarty & Waller, 2005; Santamaría et al., 2005).

Relationships between CalibrainVR-MR and academic performance measures (GPA, math, and science scores) were modest, explaining only 4–6% of the variance. However, the stronger association between CalibrainVR-MR performance and academic outcomes in STEM programs compared to non-STEM programs supports its external validity and aligns with previous research emphasizing the critical role of spatial abilities in STEM disciplines (Sorby et al., 2013; Uttal & Cohen, 2012; Veurinck & Sorby, 2019). Interestingly, no significant difference was found in CalibrainVR-MR scores between students in STEM and non-STEM programs, suggesting that students across disciplines may possess similar mental rotation abilities. However, the greater relevance of these abilities within STEM contexts suggests an opportunity to identify untapped potential among non-STEM students and encourage their participation in STEM fields. This finding aligns with Wai et al. (2009), who advocated for incorporating spatial abilities into talent searches to uncover overlooked STEM potential.

Boys and girls exhibited equal variance in their CalibrainVR-MR scores, with no significant sex differences observed. This finding is consistent with research showing that traditional sex differences in 2D-based mental rotation tasks (Voyer et al., 1995) diminish or disappear when tasks involve 3D objects (Larson et al., 1999; Neubauer et al., 2010; Parsons et al., 2004). The results underscore the potential of VR-based assessments to provide more equitable evaluations of mental rotation abilities by eliminating artificial gender gaps often introduced by traditional 2D tasks.

In conclusion, Chapter 3 underscores the promise of CalibrainVR-MR and similar VR-based tools in advancing the assessment of spatial abilities. These tools not only improve the realism and fairness of evaluations but also offer a novel way to identify STEM talent that may be overlooked by traditional assessments. By incorporating spatial ability into educational guidance and talent searches, VR-based tools can broaden participation in STEM fields, support equitable evaluation, and help students make informed educational choices.

SIMON-jr and PE Interest Fit

Chapter 4 focused on examining the role of person-environment (PE) interest fit in secondary education, using the SIMON-jr instrument. This chapter addressed a significant gap in career exploration research, which has largely focused on adults and higher education students, overlooking early adolescents despite the important academic and career decisions they face during this formative period (Seghers et al., 2019). By investigating vocational interests and PE fit among secondary students, the study aimed to assess the validity of the SIMON-jr, uncover potential sex differences in PE interest fit, and explore the relationship between PE fit and academic achievement.

The validity of the SIMON-jr was assessed through three key aspects: internal consistency, internal structure, and sex differences. The tool demonstrated good internal consistency for five RIASEC dimensions (Realistic, Investigative, Artistic, Social, and Enterprising), supporting its use in educational counseling. While the Conventional dimension exhibited comparatively lower reliability ($\alpha = .62$), it remained within acceptable limits for group-level research (Evers et al., 2010). Confirmatory factor analysis (CFA) validated the alignment of the SIMON-jr with the RIASEC framework, with significant factor loadings confirming the robustness of the items. The expected circular structure of the RIASEC model was not observed, aligning with prior research suggesting that this structure emerges later in adolescence (Tracey, 2002; Tracey & Rounds, 1993).

Sex differences in vocational interests followed established trends (Su et al., 2009), with boys reporting higher Realistic and Investigative interests and girls showing stronger

Social and Artistic preferences. Notably, boys demonstrated higher Conventional interests than girls, a reversal of typical trends observed in adults but consistent with previous findings in similar age groups (Gfrörer et al., 2022). Differential item functioning (DIF) analysis indicated minimal bias, confirming the fairness of the scales across genders.

PE interest fit analyses showed that students generally select secondary education programs that fit moderately with their interests (mean fit = .57). This value is lower than the mean fit observed in higher education (Schelfhout et al., 2019), consistent with literature suggesting that vocational interests are less differentiated in early adolescence and become more specific with age (Tracey, 2002). Additionally, girls exhibited stronger PE interest fit than boys, likely due to their more differentiated interest profiles during early adolescence (Slot et al., 2021; Sung et al., 2017).

Contrary to expectations, no significant relationship was found between PE interest fit and academic achievement, nor when analyzed separately by gender. This finding aligns with meta-analytic results from higher education studies, which show only weak correlations between interest fit and academic outcomes (de Vries et al., 2024). The broader, more general nature of secondary education programs likely contributes to this absence, as these programs are designed to prepare students for diverse career paths. This lack of specialization leads to greater variation in RIASEC profiles within programs, creating more diverse learning environments and weaker associations with academic performance.

In summary, Chapter 4 highlights the reliability and validity of the SIMON-jr for assessing vocational interests in early adolescents and emphasizes its potential as a tool for guiding students toward informed educational choices. While PE interest fit is evident in secondary education, its weaker association with academic outcomes suggests the need for further research to refine its application and explore its long-term implications for educational and career development.

Development of PE Interest Fit

Chapter 5 explored the longitudinal development of PE interest fit during early adolescence within the Flemish educational system. Acknowledging the developmental nature of vocational interests during this stage (Hoff et al., 2018), a longitudinal study design was implemented to investigate the evolution of interests and their fit with educational choices over time. Additionally, the study assessed the capacity of the SIMON-jr to effectively capture these evolving interests and support informed decision-making in early adolescence.

Contrary to expectations (Holland, 1997; Schneider, 1987), the study found no significant improvement in PE interest fit with chosen study programs over time, for either boys or girls. This challenges the assumption that adolescents progressively gravitate toward the program that best fits with their goals. Several factors may explain this finding. The general structure of secondary education, which is less specialized than higher education, may dilute the fit between interests and programs. Additionally, external influences such as parental guidance, academic pressures, and societal expectations likely play a significant role in shaping students' decisions. Alternatively, this outcome may reflect the gradual and complex process of developing a coherent self-concept and a differentiated RIASEC profile, as suggested by prior research (Slot et al., 2021).

Despite the lack of improvement, the observed level of PE interest fit in secondary education is relatively strong and remains stable over time, indicating that many students make choices broadly consistent with their vocational preferences. This finding is notable given the developmental shifts in interests during adolescence, where most vocational interest dimensions show a general decline in intensity. The stability of PE interest fit amidst these changes suggests that students' choices continue to reflect their core interests. This finding underscores the value of supporting interest-driven educational decisions in adolescence, as well as the robustness of PE interest fit as a framework for navigating these choices, even amidst the natural shifts in interest that occur during early adolescence.

In addition, students showed a decline in PE interest fit with unselected programs. Although this decline was not statistically significant for girls, additional analysis revealed that the difference between the fit with chosen and unchosen programs became increasingly pronounced over time for both boys and girls. This pattern aligns with Gottfredson's theory of circumscription and compromise (1981), which posits that as individuals clarify their career trajectories and refine their self-concept, they increasingly distance themselves from previously considered but ultimately unselected options.

In conclusion, while adolescents' vocational interests evolve, this evolution does not necessarily result in improved fit with chosen study programs. Instead, students—particularly boys—tend to narrow their focus by disengaging from excluded options. Additionally, over time, the difference between fit with chosen and unchosen programs becomes more pronounced for both boys and girls, largely due to a decrease in fit with the unchosen options. These findings highlight the complexity of vocational interest development and the interplay of individual and structural factors. The SIMON-jr instrument demonstrated utility in capturing these dynamics, reinforcing its value in guiding early adolescents' educational choices.

General Conclusions and Contributions of the Empirical Chapters

Table 1 summarizes the key findings from the empirical chapters of this dissertation. By examining cognitive abilities, vocational interests, and the fit between students' interests and educational programs, this research reveals how these factors influence students' academic paths and career goals. The development and validation of the Ch-ICAR and CalibrainVR-MR provide psychometrically sound tools for assessing cognitive abilities and spatial skills, which are crucial predictors of academic success. Additionally, the findings from the SIMON-jr expand our understanding of vocational interests, showing that PE interest fit plays a vital role in guiding educational decisions in secondary education. The studies reported in this dissertation underscore the complexity of career exploration in secondary education and suggest that a dynamic, continuous process of exploration is essential for optimizing study choice and long-term career outcomes. This aligns with career construction theory, which posits that individuals use an adaptive process to develop their careers and achieve fit between the self and the environment, ultimately supporting vocational development (Savickas, 2002, 2013).

Table 1

Summary of the Empirical Chapters

Chapter	Key research questions	Theoretical framework	Sample	Main findings
Chapter 2	How valid and reliable is the Ch-ICAR in measuring cognitive abilities in children, and how does the test compare to established intelligence tests?	Cattell-Horn-Carroll (CHC) framework (Schneider & McGrew, 2012)	-99 primary school students (pilot study) - 820 secondary school students (Study 1) - 91 secondary school students and 96 primary school students (Study 2)	- The Ch-ICAR is a reliable and valid tool for measuring children's cognitive abilities - Study 1: the Ch-ICAR explained significant variance in GPA (18%), mathematics scores (26%), and CDR scores (35%) - Study 2: the Ch-ICAR showed high correlations with the Raven's Progressive Matrices and Flemish CoVaT-CHC Basic Version, demonstrating its construct validity - The Ch-ICAR is mainly a measure of fluid intelligence (Gf), with subtests showing alignment with broad cognitive abilities, though caution is advised with certain subtests like Matrix Reasoning and Figural Analogies
Chapter 3	How valid and reliable is the CalibrainVR-MR to assess mental rotation abilities in secondary students, and is CalibrainVR-MR performance related to STEM performance?	Cattell-Horn-Carroll (CHC) framework (Schneider & McGrew, 2012)	1,223 secondary school students (aged 11-14) & a subset of 131 students who provided longitudinal academic performance data	-The final 34-item subset of CalibrainVR-MR demonstrated good internal consistency and is recommended for future use - Medium correlations were found between CalibrainVR-MR, mathematical skills, and cognitive abilities -While only modest relationships were found between CalibrainVR-MR and academic performance, stronger associations were seen in STEM programs compared to non-STEM programs, emphasizing the relevance of spatial skills in STEM - Boys and girls showed equal variance in their scores, with no significant gender differences, highlighting VR's potential to eliminate the gender gap often seen in 2D mental rotation tasks

Chapter 4	How valid and reliable is the SIMON-jr to measure vocational interests in early adolescents? To what extent are students' vocational interests in the first year of secondary education aligned with their study program choices three years; does PE interest fit predict academic achievement; are there sex differences in these patterns?	Holland’s theory of vocational interests (1997)	868 secondary school students (data collected in their 1st and 3rd year of secondary education)	-SIMON-jr is a reliable and valid tool for assessing vocational interests in early adolescents -Early adolescents displayed moderate PE interest fit (mean fit = .57), with girls showing stronger fit than boys -No significant relationship was found between PE interest fit and academic achievement
Chapter 5	How does PE interest fit evolve from grade 7 (ages 12–13) to grade 9 (ages 14–15), and are there sex differences in these patterns?	Holland’s theory of vocational interests (1997) & Gottfredson’s (1981) theory of circumscription and compromise	885 secondary school students (data collected in 3 consecutive school years)	-Students exhibit a substantial fit with their grade 9 program choice at the start of secondary education -While students’ interests evolve, the fit with their chosen study program remains stable over time -Students, particularly boys, progressively lose interest in unchosen options, indicating a narrowing of focus -The difference in fit between chosen and unchosen programs becomes increasingly pronounced over time for both boys and girls, with the decline in fit with unchosen options being a key factor -Our findings underscore the importance of incorporating vocational interest assessment tools into the career decision-making processes of early adolescents

Practical Implications

This dissertation offers valuable insights into improving study orientation processes and career exploration in secondary education. By integrating the findings into educational practice, particularly within frameworks such as the Flemish Talentcenters, schools and policymakers can create more inclusive, tailored, and effective approaches to guiding students through key educational decisions.

One significant implication of this research is the potential to empower students in identifying and utilizing their strengths through innovative tools such as CalibrainVR-MR. As a sex-neutral measure of mental rotation abilities, this tool not only offers students valuable insights into their spatial reasoning skills—skills often linked to success in STEM fields (Wai et al., 2009)—but also challenges persistent gender stereotypes that influence educational and career decisions (Veldman et al., 2021). By demonstrating that spatial reasoning abilities are distributed equally across sexes, CalibrainVR-MR could encourage more girls to explore and pursue pathways in traditionally male-dominated fields like engineering and technology. Integrating such tools into career education can actively contribute to reducing gender disparities in STEM enrollment and career representation.

This dissertation also underscores the robustness of PE interest fit as valuable tool for supporting young adolescents in their study choices, even amidst developmental changes. The findings reveal that while vocational interests may evolve, there is a strong connection between students' core interests during early adolescence and the educational pathways they eventually select. Using PE interest fit tools to support early decisions can empower students make confident choices, enhancing their engagement, satisfaction, and long-term success in their educational trajectories (de Vries et al., 2024; Hyland et al., 2022).

Finally, this dissertation reinforces the value of career exploration as a proactive, adaptive process (Lent & Brown, 2013). Frameworks like the Flemish Talentcenters are uniquely positioned to support students in reflecting on their strengths, exploring a range of possibilities, and making well-informed, thoughtful decisions about their educational and career pathways. By encouraging active self- and environmental exploration (Hughes et al., 2016; Kuijpers & Scheerens, 2006), the Talentcenter framework helps students build career decision self-efficacy and intrinsic motivation, equipping them with the skills needed for lifelong career management (Jiang et al., 2019).

Future Research

This dissertation documented the development, validation, and application of innovative instruments designed to enhance the study choice process in secondary education, tailored to the Flemish educational system. While the findings offer a strong foundation, there are numerous opportunities for future research to refine these instruments and explore broader implications of exploration tools like the Talentcenter.

One area for further research is ensuring fairness. To make educational opportunities truly accessible to all students, it is essential to ensure that the instruments validated in this dissertation remain unbiased and do not disadvantage any specific groups. While our research found no evidence of sex bias and took socioeconomic status (SES) into account, ongoing development should continue to monitor for such biases, ensuring that the tools remain equitable and inclusive. Additionally, this dissertation underscored the broader role of social inequalities in education, such as disparities in access to critical information and biases in teacher recommendations (Ball, 2003; Draelants, 2014; Seghers et al., 2019). Future research could adopt a longitudinal approach to examine whether the feedback provided by the Talentcenters helps reduce these disparities. Investigating whether the Talentcenters support students from underrepresented or disadvantaged groups in making informed and equitable study choices could yield critical insights into their role in addressing potential inequities in education (OECD, 2024; Seghers et al., 2019).

Further validation and refinement of the instruments developed in this dissertation—SIMON-jr, CalibrainVR-MR, and Ch-ICAR—offer another important direction for future research. For example, while the full Ch-ICAR demonstrated strong psychometric properties, we recommend against interpreting its subtests independently at this stage. Future studies should explore whether subtests like Matrix Reasoning and Figural Analogies can provide reliable and meaningful standalone results, requiring replication of findings and targeted improvements to these components. Moreover, it is essential to establish age- or grade-based norms for the Ch-ICAR, as cognitive abilities undergo dynamic changes during childhood and adolescence (Flanagan & McDonough, 2018). Such norms would allow for more precise and developmentally appropriate interpretations of results. Similarly, while CalibrainVR-MR demonstrated promise as a sex-neutral measure of mental rotation abilities, its potential to encourage greater participation of girls in STEM pathways deserves further exploration. Longitudinal research could examine whether exposure to this tool builds confidence in spatial abilities and influences subsequent study choices and career aspirations, contributing to greater gender equity in STEM fields.

Research into PE interest fit, a key focus of this dissertation, is another area where future research could broaden understanding. Investigating specific subpopulations, such as students who switch programs or drop out, could provide valuable insights into the consequences of misaligned interests, including disengagement, academic struggles, or dissatisfaction with chosen programs. By better understanding these dynamics, interventions could be developed to support at-risk students and guide them toward educational paths better aligned with their interests and strengths.

In addition, the Talentcenter framework itself offers several opportunities for further development and research. First, expanding the range of competencies assessed in the Talentcenter is a promising direction. While this dissertation addressed attributes critical to study choice and academic achievement, future research could explore the predictive validity of additional variables, such as problem-solving skills (Greiff et al., 2014), emotional intelligence (MacCann et al., 2020), or computational thinking (Román-González et al., 2017). These variables are increasingly recognized as essential skills in a rapidly changing society and could further diversify the range of measured competencies, making the Talentcenter assessment more comprehensive. Emotional intelligence, for instance, is a significant predictor of academic achievement, with stronger associations observed in social and humanities fields (MacCann et al., 2020). Conversely, problem-solving skills and computational thinking are more closely linked to performance in STEM disciplines (Greiff et al., 2014; Román-González et al., 2017). Including these variables in future research could not only enhance the breadth of the Talentcenter but also provide theoretical insights by examining their relationship with study choices in secondary education. This would expand our understanding of how diverse skills influence academic orientation and success across different fields. Additionally, employing more ecologically valid measurement approaches, such as virtual reality (VR) environments that simulate real-world challenges, could enhance the tools' relevance and applicability (Parsons, 2015).

Second, while the Talentcenter's feedback model, which draws on data from a large longitudinal study, offers substantial utility, continued longitudinal data collection would enable more dynamic and precise feedback tailored to students' evolving needs. Integrating additional variables and analyzing long-term outcomes could improve the model's ability to deliver actionable, personalized recommendations. These insights could increase students' engagement in exploratory behaviors, such as self- and environmental exploration, which are crucial for effective career planning and progression (e.g., Lent et al., 2016; Lent et al., 2017) and in turn may help reduce rates of reorientation and dropout (Fall & Roberts, 2012; Janosz et al., 2008; Rumberger & Rotermund, 2012).

Finally, the personalized feedback report generated by the Talentcenter presents an opportunity for further investigation. Future studies could assess how students perceive, trust, and engage with the feedback (Carless & Boud, 2018). Do they understand the recommendations? Do they explore suggested pathways? Do they use the feedback as a meaningful tool in their study choice process? Formal user feedback research could provide actionable insights, helping to optimize the content and format of the reports to better meet the diverse needs of students (Van Boekel et al., 2023).

By pursuing these avenues of research, both the instruments developed in this dissertation and the broader Talentcenter framework can be further refined, enhancing their utility as tools for guiding students toward informed, equitable, and fulfilling educational choices.

Conclusion

This dissertation provides significant empirical insights into enhancing study orientation and career exploration processes in secondary education. Across its studies, the research validated innovative tools—CalibrainVR-MR, SIMON-jr, and Ch-ICAR—that assess critical abilities and interests linked to educational decisions and academic performance. These tools were shown to offer reliable, developmentally appropriate assessments while promoting inclusivity. In addition, the findings highlight the potential of exploration tools like the Flemish Talentcenter, which provide a comprehensive framework to support students in making informed study choices. By encouraging self-reflection on strengths and preferences, resources like the Talentcenter empower students to select study environments that match their personal interests and strengths. Such approach holds particular promise for addressing systemic challenges, such as advancing gender equity in STEM and mitigating the influence of social inequalities on study orientation. In sum, this dissertation underscores the role of structured career exploration in empowering students to navigate complex educational and career landscapes with confidence and clarity.

References

Ball, S. J. (2003). Class strategies and the educational marketplace: The middle-classes and social advantage. London: Routledge

Carless, D. & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43,* 1315-1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, L. L. (2022). Spatial Ability and Mathematics Achievement in Deaf Children: The Mediating Role of Processing Speed and Intelligence. *Journal of Developmental and Physical Disabilities, 34,* 399-415. https://doi.org/10.1007/s10882-021-09805-w

Condon, D., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Dawis, R. V. (2005). The Minnesota Theory of Work Adjustment. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (pp. 3-23). Hoboken, NJ: John Wiley & Sons Inc.

de Vries, N., Meeter, M., & Huizinga, M. (2024). Does interest fit between student and study program lead to better outcomes? A meta-analysis of vocational interest congruence as predictor for academic success. *Educational Research Review, 44,* 100619. https://doi.org/10.1016/j.edurev.2024.100619

Draelants, H. (2014). Des héritiers aux initiés? Note sur les nouvelles modalités de la reproduction sociale par l'école. *Social Science Information, 53,* 403–432. https://doi.org/ 10.1177/0539018414522893

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN Review System for Evaluating Test Quality.* Boom Uitgevers Amsterdam. Retrieved from https://psynip.nl/wp-content/uploads/2022/05/COTAN-review-system-for-evaluating-test-quality.pdf

Fall, M. A., & Roberts, G. (2012). High School Dropouts: Interactions between Social Context, Self-Perceptions, School Engagement, and Student Dropout. *Journal of Adolescence, 35,* 787–798. https://doi.org/10.1016/j.adolescence.2011.11.004.

Flanagan, D. P., & McDonough, E. M. (2018). *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (Fourth Edition). Guilford Press.

Gfrörer, T., Stoll, G., Rieger, S., Trautwein, U., & Nagengast, B. (2022). The development of vocational interests in early adolescence: Stability, change, and state-trait components. *European Journal of Personality, 36,* 942. https://doi.org/10.1177/08902070211035630

Goosen, K., Boone, S., Dehertogh, B., Kavadias, D., Mahieu, P., Van Avermaet, P., … Van Kerckhove, C. (2017). *Is dat iets voor mij juf? Leerlingen versterken in het keuzeproces van basis naar secundair.* Leuven: LannooCampus.

Gottfredson, L. S. (1981). Circumscription and compromise: A developmental theory of occupational aspirations. *Journal of Counseling Psychology, 28,* 545–579. https://doi.org/10.1037/0022-0167.28.6.545

Greiff, S., Kretzschmar, A., Leutner, D. (2014). Problem Solving in Educational Psychology. *Editorial Material, 28,* 161-166. https://doi.org/ 10.1024/1010-0652/a000140

Hegarty, M., & Waller, D. A. (2005). Individual Differences in Spatial Abilities. In P. Shah & A. Miyake (Eds.), The Cambridge Handbook of Visuospatial Thinking (121–170). Cambridge University Press.

Hirschi, A. (2009). Career adaptability development in adolescence: Multiple predictors and effect on sense of power and life satisfaction. *Journal of Vocational Behavior, 74,* 145–155. https://doi.org/10.1016/j.jvb.2009.01.002.

Hoff, K. A., Briley, D. A., Wee, C. J. M., & Rounds, J. (2018). Normative changes in interests from adolescence to adulthood: A meta-analysis of longitudinal studies. *Psychological Bulletin, 144,* 426–451. https://doi.org/ 10.1037/bul0000140

Holland, J. L. (1997). *Making vocational choices: A theory of vocational personalities and work environments* (3rd ed.). Odessa, FL, US: Psychological Assessment Resources.

Hughes, D., Mann, A., Barnes, S.-A., Baldauf, B., & McKeown, R. (2016). *Careers education: International literature review.* Warwick Institute for Employer Research and Education and Employers Research. https://www.educationandemployers.org/wp-content/uploads/2016/07/Careers-review.pdf

Hyland, W. E., Hoff, K. A., & Rounds, J. (2022). Interest–ability profiles: An integrative approach to knowledge acquisition. *Journal of Intelligence, 10,* 43. https://doi.org/10.3390/jintelligence10030043

Janosz, M., I. Archambault, J. Morizot, & Pagani, L. S. (2008). School engagement trajectories and their differential predictive relations to dropout. *Journal of Social Issues, 64,* 21–40. https://doi.org/10.1111/josi.2008.64.issue-1

Jiang, Z., Newman, A., Le, H., Presbitero, A., & Zheng, C. N. (2019). Career exploration: A review and future research agenda. *Journal of Vocational Behavior, 110,* 338-356. https://doi.org/ 10.1016/j.jvb.2018.08.008

Kuijpers, M., & Scheerens, J. (2006). Career competencies for the modern career. Journal of *Career Development, 32,* 303–319. https://doi.org/10.1177/ 0894845305283006

Larson, P., Rizzo, A. A., Buckwalter, J. G., & Van Roo. A. (1999). Gender issues in the use of virtual environments. *Cyberpsychology & Behavior, 2,* 113–123. https://doi.org/10.1089/cpb.1999.2.113. PMID: 19178247

Lent, R. W., & Brown, S. D. (2013). Social cognitive model of career self-management: Toward a unifying view of adaptive career behavior across the life span. *Journal of Counseling Psychology, 60,* 557–568. https://doi.org/10.1037/a0033446.

Lent, R. W., Ezeofor, I., Morrison, M. A., Penn, L. T., & Ireland, G. W. (2016). Applying the social cognitive model of career self-management to career exploration and decision-making. *Journal of Vocational Behavior, 93,* 47–57. https://doi.org/10.1016/j.jvb.2015.12.007.

Lent, R. W., Ireland, G. W., Penn, L. T., Morris, T. R., & Sappington, R. (2017). Sources of self-efficacy and outcome expectations for career exploration and decisionmaking: A test of the social cognitive model of career self-management. *Journal of Vocational Behavior, 99,* 107–117. https://doi.org/10.1016/j.jvb.2017.01.002.

MacCann, C., Jiang, Y. X., Brown, L. E. R., Double, K. S., Bucich, M., Minbashian, A. (2020). Emotional Intelligence Predicts Academic Performance: A Meta-Analysis. *Psychological Bulletin, 146,* 150-186. https://doi.org/10.1037/bul0000219

Magez, W., Tierens, M., Van Huynegem, J., Van Parijs, K., Decaluwé, V., & Bos, A. (2015). *CoVaT-CHC Basisversie.* Antwerpen: Psychodiagnostisch Centrum (PDC), Thomas More Hogeschool

McLeod, J. W. H., & McCrimmon, A. W. (2021). Test Review: Raven's 2 Progressive Matrices, Clinical Edition (Raven's 2). *Journal of Psychoeducational Assessment, 39*, 388–392. https://doi.org/10.1177/0734282920958220

Neubauer, A. C., Bergner, S., & Schatz, M. (2010). Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence, 38,* 529–539. https://doi.org/10.1016/j.intell.2010.06.001

OECD. (2024). *Education at a glance 2024*. Retrieved from https://www.oecd.org/en/publications/education-at-a-glance-2024_c00cad36-en.html

Parsons, T. D. (2015). Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Frontiers in Human Neuroscience, 9,* 660. https://doi.org/10.3389/fnhum.2015.00660

Parsons, T. D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J. G., & Rizzo, A. A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia, 42,* 555–562. https://doi.org/10.1016/j.neuropsychologia.2003.08.014

Revelle, W., Dworak, E., & Condon, D. (2020). Cognitive ability in everyday life: The utility of open-source measures. *Current Directions in Psychological Science*, *29*(4), 358–363. https://doi.org/10.1177/0963721420922178

Román-González, M., Pérez-González, J. C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. *Computers in Human Behavior, 72,* 678-691. https://doi.org/10.1016/j.chb.2016.08.047

Rumberger, R. W., & Rotermund, S. (2012). The relationship between engagement and high school dropout. In *Handbook of Research on Student Engagement*, edited by S. L. Christenson, A. L. Reschly, and C. Wylie, 491–513. New York, NY: Springer.

Santamaría, P., Arribas, D., Pereña, J., & Seisdedos, N. (2005). *EFAI, Evaluación Factorial de las Aptitudes Intelectuales.* Madrid: Departamento I+D TEA Ediciones.

Savickas, M. L. (2002). Career construction: A developmental theory of vocational behavior. In D. Brown (Ed.). *Career choice and development* (pp. 255–311). (4th ed.). San Francisco, CA: Jossey-Bass.

Savickas, M. L. (2013). Career construction theory and practice. In S. D. Brown, & R. W. Lent (Eds.). *Career development and counseling: Putting theory and research to work* (pp. 147–183). (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Schelfhout, S., Wille, B., Fonteyne, L., Roels, E., De Fruyt, F., & Duyck, W. (2019). The effects of vocational interest on study results: Student Person – Environment fit and program interest diversity. *PLoS One, 14*, e0214618, https://doi.org/10.1371/journal.pone.0214618

Schneider, B. (1987). The people make the place. *Personnel Psychology, 40,* 437-453. doi:10.1111/j.1744-6570.1987.tb00609.x

Seghers, M., Boone, S., & Van Avermaet, P. (2019). Classed patterns in the course and outcome of parent-teacher interactions regarding educational decision-making. *Educational Review, 73,* 417-435. https://doi.org/10.1080/00131911.2019.1662771

Slot, E. M., Bronkhorst, L. H., Akkerman, S. F., & Wubbels, T. (2021). Vocational interest profiles in secondary school: Accounting for multiplicity and exploring associations with future-oriented choices. *Journal of Educational Psychology, 113,* 1059-1071. https://doi.org/10.1037/edu0000475

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences, 26,* 20–29. https://doi.org/10.1016/j.lindif.2013.03.010

Sorby, S., Veurink, N., & Streiner, S. (2018). Does spatial skills instruction improve STEM outcomes? The answer is 'yes.' *Learning and Individual Differences, 67,* 209–222. https://doi.org/10.1016/j.lindif.2018.09.001

Su R., Rounds J., Armstrong P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological bulletin, 135,* 859-884. https://doi.org/10.1037/a0017364

Sung, Y. T., Cheng, Y. W., & Hsueh, J. H. (2017). Identifying the career-interest profiles of junior-high-school students through latent profile analysis. *The Journal of Psychology, 151,* 229 –246. http://doi .org/10.1080/00223980.2016.1261076

Tracey, T. J. G. (2002). Development of Interests and Competency Beliefs: A 1-Year Longitudinal Study of Fifth- to Eighth-Grade Students Using the ICA–R and Structural Equation Modeling. *Journal of Counseling Psychology, 49* 148-163. https://doi.org/10.1037//0022-0167.49.2.148

Tracey, T. J. G., & Rounds, J. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin, 113*, 229-246. https://doi.org/10.1037/0033-2909.113.2.229

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how? In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 57, pp. 147-181). Oxford, UK: Academic Press.

Van Boekel, M., Hufnagle, A. S., Weisen, S., & Troy, A. (2023). The feedback I want versus the feedback I need: Investigating students' perceptions of feedback. *Psychology in the Schools, 60,* 3389-3402. https://doi.org/10.1002/pits.22928

Veldman, J., Van Laar, C., Thoman, D. B., & Van Soom, C. (2021). "Where will I belong more?": The role of belonging comparisons between STEM fields in high school girls' STEM interest. *Social Psychology of Education, 24*, 1363–1387. https://doi.org/10.1007/s11218-021-09663-6

Veurink, N. L., & Sorby, S. A. (2019). Longitudinal study of the impact of requiring training for students with initially weak spatial skills. *European Journal of Engineering Education, 44,* 153–163. https://doi.org/10.1080/03043797.2017.1390547

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250–270. https://doi.org/10.1037/0033-2909.117.2.250. PMID: 7724690.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835. https://doi.org/10.1037/a0016127

Young, S., & Keith, T. (2020). An Examination of the Convergent Validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment*, *38*(8), 1052–1059. https://doi.org/10.1177/0734282920943455

Appendices Chapter 2

Appendix 2A

Table 2A1

Specimen of Each Item Type of the Ch-ICAR (Loe et al., 2018; The International Cognitive Ability Resource Team, 2014)

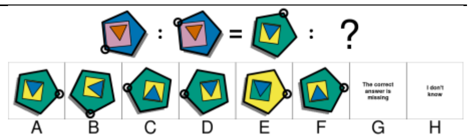
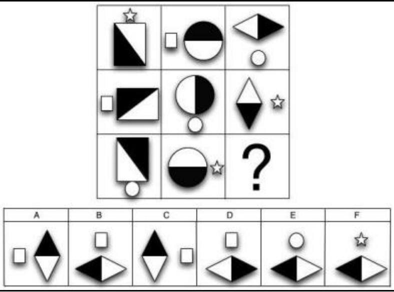
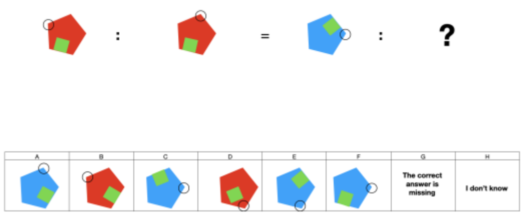
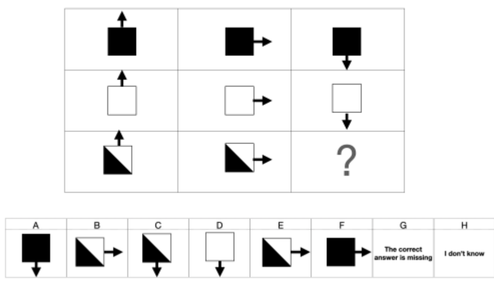
Verbal Reasoning	If the day after tomorrow is two days before Thursday then what day is it today? (1) Friday (2) Monday (3) Wednesday (4) Saturday (5) Tuesday (6) Sunday (7) None of these (8) I don't know	Figural Analogies	
Matrix Reasoning		Number Series	34,150,35,160,36,170, …

Table 2A2

Example Items of the Subtest Figural Analogies and Matrix Reasoning

Figural Analogies	Matrix Reasoning
	

Appendix 2B

Table 2B1

Overview Test Statistics Verbal Reasoning Items (Pilot Study, N = 46)

Cronbach's alpha full item set (16 items): .50
Cronbach's alpha final item selection (8 items): .65

ICAR item number	PC	Mean response time (in sec)	PC boys	PC girls
VR18	.91	24	.91	.92
VR31	.56	42	.52	.58
VR39	.77	23	.90	.65
VR11	.12	19	.10	.13
VR14	.42	37	.50	.35
VR17	.37	41	.25	.48
VR42	.42	18	.40	.43
VR04	.10	22	.16	.04
VR19	.52	35	.53	.52
VR32	.43	39	.53	.35
VR09	.03	22	.00	.04
VR16	.21	37	.35	.09
VR36	.03	18	.06	.00
VR23	.26	19	.41	.14
VR26	.00	17	.00	.00
VR13	.15	40	.18	.14

Note. The numbers of the items refer to the numeration in the ICAR database (The International Cognitive Ability Resource Team, 2014). PC = proportion of correct responses in the full sample; PC boys = proportion of correct responses in boys subsample; PC girls = proportion of correct responses in girls subsample.

Table 2B2

Overview Test Statistics Matrix Reasoning Items (Pilot Study, N = 66)

Cronbach's alpha full item set (11 items): .63
Cronbach's alpha final item selection (8 items): .69

ICAR item number	PC	Mean response time (in sec)	PC boys	PC girls
fig12043	.52	62	.64	.41
fig12044	.52	41	.56	.49
fig12047	.36	36	.41	.33
fig12046	.30	33	.33	.28
fig12053	.26	39	.30	.23
fig12045	.29	28	.30	.28
fig12048	.23	34	.33	.15

fig12054	.17	31	.22	.13
fig12056	.31	27	.33	.30
fig12055	.16	29	.22	.11
fig12050	.13	26	.11	.14

Note. The numbers of the items refer to the numeration in the ICAR database (The International Cognitive Ability Resource Team, 2014). PC = proportion of correct responses in the full sample; PC boys = proportion of correct responses in boys subsample; PC girls = proportion of correct responses in girls subsample.

Table 2B3

Overview Test Statistics Number Series Items (Pilot Study, N = 71)

Cronbach's alpha full item set (23 items): .88
Cronbach's alpha final item selection (8 items): .85

ICAR item number	PC	Mean response time (in sec)	PC boys	PC girls
8	.83	33	.83	.83
6	.90	20	.90	.90
2	.66	34	.79	.56
7	.89	18	.86	.90
1	.63	28	.79	.51
3	.51	35	.66	.39
4	.63	28	.79	.51
5	.52	28	.61	.44
10	.44	33	.52	.39
12	.49	28	.62	.40
9	.20	46	.19	.20
25	.22	45	.19	.24
19	.25	41	.23	.26
26	.32	41	.31	.33
27	.37	45	.46	.30
11	.09	56	.13	.06
33	.00	44	.00	.00
42	.09	49	.20	.03
13	.08	37	.05	.09
16	.10	39	.05	.12
44	.08	34	.11	.06
40	.06	38	.11	.03
28	.04	22	.06	.03

Note. The numbers of the items refer to the numeration in the ICAR database (The International Cognitive Ability Resource Team, 2014). PC = proportion of correct responses in the full sample; PC boys = proportion of correct responses in boys subsample; PC girls = proportion of correct responses in girls subsample.

Table 2B4

Overview Test Statistics Figural Analogies Items (Pilot Study, N = 56)

Cronbach's alpha full item set (20 items): .77
Cronbach's alpha final item selection (7 items): .83

ICAR item number	**PC**	**Mean response time (in sec)**	**PC boys**	**PC girls**
q18005	.36	68	.30	.41
q18019	.38	35	.31	.45
q18020	.44	28	.36	.52
q18006	.51	25	.52	.50
q18002	.26	24	.16	.36
q18001	.25	20	.16	.32
q18004	.25	36	.24	.25
q18009	.26	24	.13	.37
q18010	.31	19	.17	.42
q18003	.27	22	.13	.38
q18012	.29	28	.39	.19
q18011	.19	31	.26	.12
q18014	.15	20	.14	.16
q18018	.13	21	.14	.12
q18013	.33	18	.29	.38
q18017	.07	21	.05	.08
q18016	.24	19	.19	.29
q18015	.22	21	.19	.25
q18008	.05	16	.05	.04
q18007	.09	20	.15	.04

Note. The numbers of the items refer to the numeration in the ICAR database (The International Cognitive Ability Resource Team, 2014). PC = proportion of correct responses in the full sample; PC boys = proportion of correct responses in boys subsample; PC girls = proportion of correct responses in girls subsample.

Appendix 2C

Table 2C1

Included ICAR Items in the Children's ICAR

Subtest	**ICAR item number**
Verbal Reasoning	VR39, VR31, VR19, VR14, VR32, VR16, VR11, VR04
Matrix Reasoning	12044, 12043, 12056, 12046, 12053, 12048, 12055, 12050
Number Series	8, 5, 3, 12, 10, 27, 25, 13
Figural Analogies	q18005, q18020, q18019, q18006, q18015, q18009, q18016

Note. The numbers of the items refer to the numeration in the ICAR database (The International Cognitive Ability Resource Team, 2014)

Appendix 2D

Table 2D1

Item Fit Final Hierarchical IRT Model (Study 1, N = 820)

Item	S_χ^2 Statistic	df	RMSEA	FDR adjusted p-value
MR1	12.5	15	.000	.719
MR2	24.4	13	.033	.105
MR3	20.2	14	.023	.258
MR4	12.9	14	.000	.632
MR6	11.8	15	.000	.719
MR7	8.4	15	.000	.906
VR1	15.3	13	.015	.433
VR2	23.9	11	.038	.060
VR3	16.7	13	.019	.382
VR4	28.1	12	.041	.048
VR5	35.0	14	.043	.020
VR6	17.0	14	.016	.431
VR7	16.5	14	.015	.433
VR8	17.2	13	.020	.363
NS1	19.2	12	.027	.215
NS2	26.8	13	.036	.060
NS3	10.9	10	.011	.509
NS4	35.9	10	.056	.002
NS5	16.6	9	.032	.169
NS7	14.0	13	.009	.509
NS8	12.5	13	.000	.601
FA1	11.2	11	.005	.546
FA2	17.7	11	.027	.215
FA3	19.3	11	.030	.169
FA4	18.7	12	.026	.215
FA5	11.1	14	.000	.719
FA6	30.0	14	.037	.052

Appendix 2E

Table 2E1

Correlations Between Academic Performance and SES Indicators (Study 1, N = 820)

	1.	2.	3.	4.	5.	6.	7.
1. GPA	1						
2. Mathematics	.86**	1					
3. CDR	.37**	.40**	1				
4. Home language	-.11**	-.10**	-.14**	1			
5. Bursary	-.27**	-.27**	-.23**	.20**	1		
6. Education parent 1	-.02	-.02	-.15**	.29**	.24**	1	
7. Education parent 2	-.11**	-.10*	-.18**	.29**	.28**	.46**	1

Note. *<.05; **<.01.

Appendix 2F

Table 2F1

Internal Consistency of Ch-ICAR (sub)tests in CoVaT Sample (Study 2, N = 96)

(sub)test	McDonald's omega (ω)	Cronbach's alpha (α)
Verbal Reasoning	.60	.58
Matrix Reasoning	.52	.49
Number Series	.76	.74
Figural Analogies	.77	.73
Full Ch-ICAR	.84	.83

Table 2F2

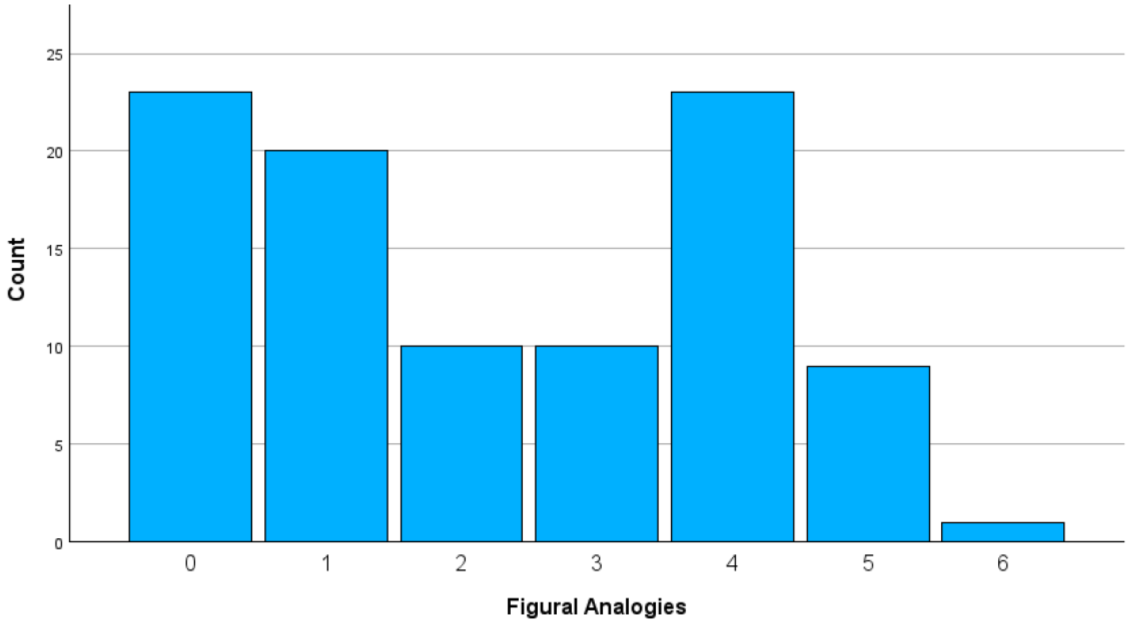
Internal Consistency of Ch-ICAR (sub)tests in RPM Sample (Study 2, N = 91)

(sub)test	McDonald's omega (ω)	Cronbach's alpha (α)
Verbal Reasoning	.66	.66
Matrix Reasoning	.31	.31
Number Series	.77	.75
Figural Analogies	.75	.70
Full Ch-ICAR	.81	.81

Appendix 2G

Figure 2G1

Distribution of Figural Analogies Scores (Study 2, CoVaT sample, N = 96)



Appendix 2H

Table 2H1

Proportion of Correct Responses per Item (Study 1 & Study 2)

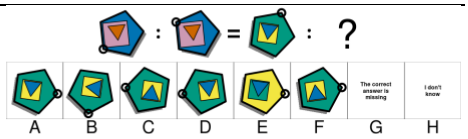
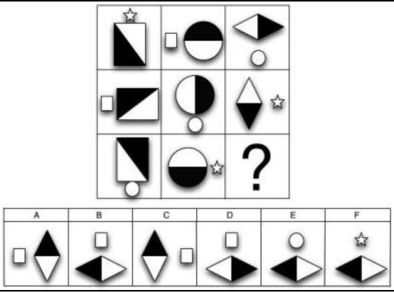
Item	PC *(SD)* Study 1, *N* = 820	PC (*SD*) Study 2 CoVaT sample, N = 96	PC (*SD*) Study 2 RPM sample, N = 91
VR1	.83 *(.38)*	.89 *(.32)*	.78 *(.42)*
VR2	.73 *(.44)*	.84 *(.37)*	.87 *(.34)*
VR3	.48 *(.50)*	.53 *(.50)*	.52 *(.50)*
VR4	.38 *(.49)*	.44 *(.50)*	.41 *(.49)*
VR5	.43 *(.50)*	.43 *(.50)*	.47 *(.50)*
VR6	.20 *(.40)*	.21 *(.41)*	.35 *(.48)*
VR7	.20 *(.40)*	.29 *(.46)*	.33 *(.47)*
VR8	.24 *(.43)*	.30 *(.46)*	.25 *(.44)*
MR1	.38 *(.49)*	.44 *(.50)*	.47 *(.50)*
MR2	.49 *(.50)*	.73 *(.45)*	.71 *(.45)*
MR3	.24 *(.43)*	.31 *(.47)*	.32 *(.47)*
MR4	.30 *(.46)*	.35 *(.48)*	.33 *(.47)*
MR5	.30 *(.46)*	.49 *(.50)*	.32 *(.47)*
MR6	.29 *(.45)*	.36 *(.48)*	.46 *(.50)*
MR7	.17 *(.37)*	.35 *(.48)*	.21 *(.41)*
MR8	.14 *(.34)*	.18 *(.38)*	.18 *(.38)*
NS1	.82 *(.38)*	.87 *(.33)*	.89 *(.31)*
NS2	.51 *(.50)*	.56 *(.50)*	.55 *(.50)*
NS3	.56 *(.50)*	.67 *(.47)*	.70 *(.46)*
NS4	.42 *(.49)*	.44 *(.50)*	.55 *(.50)*
NS5	.46 *(.50)*	.59 *(.49)*	.68 *(.47)*
NS6	.19 *(.39)*	.20 *(.40)*	.20 *(.40)*
NS7	.25 *(.43)*	.21 *(.41)*	.42 *(.50)*
NS8	.17 *(.38)*	.24 *(.43)*	.33 *(.47)*
FA1	.45 *(.50)*	.42 *(.50)*	.59 *(.49)*
FA2	.49 *(.50)*	.51 *(.50)*	.67 *(.47)*
FA3	.48 *(.50)*	.49 *(.50)*	.68 *(.47)*
FA4	.52 *(.50)*	.49 *(.50)*	.66 *(.48)*
FA5	.08 *(.27)*	.11 *(.32)*	.13 *(.34)*
FA6	.20 *(.40)*	.20 *(.40)*	.24 *(.43)*
FA7	.15 *(.36)*	.21 *(.41)*	.26 (.44)

Appendices Chapter 3

Appendix 3A

Table 3A1

Example of Each Item Type of the Ch-ICAR (Dutry et al., 2025; The International Cognitive Ability Resource Team, 2014)

Verbal Reasoning	If the day after tomorrow is two days before Thursday then what day is it today? (1) Friday (2) Monday (3) Wednesday (4) Saturday (5) Tuesday (6) Sunday (7) None of these (8) I don't know	Figural Analogies	
Matrix Reasoning		Number Series	34,150,35,160,36,170, …

Appendix 3B

Table 3B1

Overview of Item Parameters

Item n°	Condition	Rotation	Mirrored	*N* Valid responses	PC	Mean RTc	Cronbach's alpha if item deleted	Removed in final itemset
1	1	17	False	1198	.97	9.34	.62	No
2	1	14	False	1115	.95	7.06	.62	No
3	1	10	True	1207	.34	7.90	.60	No
4	1	30	False	1162	.91	6.92	.62	No
5	1	16	True	1211	.27	8.18	.58	No
6	1	22	False	1095	.86	7.47	.61	No
7	1	29	True	1201	.47	7.52	.60	No
8	1	22	True	1149	.63	5.95	.60	No
9	1	24	False	1074	.86	6.62	.62	No
10	1	29	False	1093	.90	6.63	.61	No
11	1	26	False	1159	.48	9.01	.61	No
12	1	15	True	1184	.36	7.77	.59	No
13	1	17	True	1191	.41	7.06	.58	No
14	1	29	False	1074	.87	6.46	.61	No
15	1	16	False	1056	.89	6.43	.62	No
16	1	13	True	1168	.50	7.20	.59	No
17	1	15	False	950	.91	5.48	.61	No
18	1	11	False	1104	.93	5.89	.61	No
19	1	14	False	1064	.79	6.55	.61	No
20	1	26	False	1081	.73	6.97	.61	No
21	2	159	True	1208	.43	9.11	.60	No
22	2	159	True	1145	.59	6.73	.60	No
23	2	170	True	1165	.43	6.87	.61	No
24	2	162	False	1144	.37	8.49	.63	Yes
25	2	150	True	1180	.22	7.85	.59	No
26	2	166	True	1160	.31	7.49	.60	No
27	2	157	True	1135	.38	6.45	.60	No
28	2	155	True	1125	.53	6.94	.61	No
29	2	158	False	1148	.16	7.13	.63	Yes
30	2	153	False	1053	.45	6.55	.63	Yes
31	2	153	False	1043	.53	6.91	.63	Yes
32	2	157	True	1189	.22	7.37	.60	No
33	2	157	False	1086	.67	6.61	.63	Yes
34	2	168	True	1106	.56	6.46	.63	Yes
35	2	152	True	1183	.23	6.96	.61	No

36	2	156	True	1147	.32	7.73	.62	No
37	2	154	True	1113	.33	6.24	.61	No
38	2	159	True	1090	.41	6.76	.61	No
39	2	169	True	1112	.29	6.14	.60	No
40	2	167	True	1047	.55	6.14	.60	No

Note. RTc = response time for correct responses in seconds, PC = proportion of correct responses.

Appendix 3C

Table 3C1

Correlations Between Academic Achievement and SES Indicators

	1.	2.	3.	4.	5.	6.
1. GPA	1					
2. Mathematics	.85**	1				
3. Sciences	.83**	.71**	1			
4. Home language	-.11	-.09	-.08	1		
5. Bursary	-.07	-.05	-.06	.10	1	
6. Education Parent 2	-.12	-.15	-.22*	-.03	.21*	1

Note. $^{**}p < .01$; $^{*}p < .05$; Education Parent 1 is not included in the correlation matrix due to its lack of variance among participants.

Appendix 3D

Table 3D1

Results of Differential Item Functioning Analysis

Item n°	Mantel-Haenszel χ^2	*p*	Effect size (DeltaMH)	DIF detected
1	.23	.63	.67 (A)	No
2	.00	.99	-.01 (A)	No
3	.45	.50	-.26 (A)	No
4	.01	.94	.15 (A)	No
5	2.32	.13	-.57 (A)	No
6	.06	.80	.07 (A)	No
7	4.99	.03	-.71 (A)	Yes
8	.20	.65	.16 (A)	No
9	1.29	.26	.28 (A)	No
10	.09	.76	.27 (A)	No
11	5.02	.03	-.57 (A)	Yes
12	.00	.99	.05 (A)	No
13	5.07	.02	-.79 (A)	Yes
14	2.49	.11	-.42 (A)	No
15	.02	.88	.41 (A)	No
16	.63	.43	-.16 (A)	No
17	.00	.96	-.07 (A)	No
18	1.35	.24	-.59 (A)	No
19	.52	.47	-.21 (A)	No
20	.00	.99	-.07 (A)	No
21	.60	.44	.33 (A)	No
22	3.73	.05	.55 (A)	No
23	.00	.96	.08 (A)	No
25	1.14	.29	.38 (A)	No
26	4.40	.04	.71 (A)	Yes
27	.78	.38	.23 (A)	No
28	.00	.98	.09 (A)	No
32	.12	.73	-.16 (A)	No
35	1.89	.17	-.48 (A)	No
36	.18	.67	-.07 (A)	No
37	5.86	.02	.69 (A)	Yes
38	1.34	.25	.35 (A)	No
39	9.58	.00	-.92 (A)	Yes
40	.40	.53	.25 (A)	No

Note. Effect size code A = negligible effect

Appendices Chapter 4

Appendix 4A

Items of the SIMON-jr

1. Een fiets herstellen
   (*repair a bike*)
2. Een filmpje over wetenschap bekijken (over techniek, natuur of de mens)
   (watch a video about science (about technology, nature or humans)
3. Toneelkostuums maken
   (make theater costumes)
4. Iemand troosten
   (comfort someone)
5. Mensen overtuigen om een huis te kopen
   (convince people to buy a house)
6. Controleren of mensen hun belastingen betalen
   (check if people are paying their taxes)
7. Een ijzeren tuinpoort maken
   (make an iron garden gate)
8. Werken in een laboratorium
   (work in a laboratory)
9. Een kunstwerk maken
   (create a work of art)
10. Mensen goede raad geven
    (give people good advice)
11. Een team leiden
    (lead a team)
12. Lonen berekenen
    (calculate salaries)
13. De motor van een auto herstellen
    (repair a car engine)
14. Leren over planeten
    (learn about planets)
15. Een schilderij maken
    (paint a painting)
16. Praten met mensen
    (talk to people)
17. De baas zijn
    (be the boss)
18. Controleren of de wet gevolgd wordt
    (check if the law is being followed)
19. Een tafel uit hout maken
    (make a table out of wood)
20. Mensen aan het lachen brengen
    (make people laugh)
21. Een bedrijf leiden
    (run a company)
22. Documenten opstellen, zoals een contract of een testament
    (draft documents, such as a contract or a will)

23. Mensen met problemen helpen
    (help people with problems)
24. Onderhandelen over lonen
    (negotiate wages)
25. Een knuffel geven
    (give a hug)
26. Onderhandelen over de prijs van een huis
    (negotiate the price of a house)
27. Automechanieker
    (auo mechanic)
28. Wetenschapper
    (scientist)
29. Kunstschilder
    (painter in art)
30. Rechter
    (judge)
31. Elektricien
    (electrician)
32. Onderzoeker
    (researcher)
33. Fotograaf
    (photographer)
34. Politiecommissaris
    (police commissioner)
35. Laborant
    (lab technician)
36. Artiest
    (artist)

Scoring:

Percentage of "yes" responses selected for the items corresponding to the specific interest dimension:

R = items 1, 7, 13, 19, 27, 31

I = items 2, 8, 14, 28, 32, 35

A = items 3, 9, 15, 29, 33, 36

S = items 4, 10, 16, 20, 23, 25

E = items 5, 11, 17, 21, 24, 26

C = items 6, 12, 18, 22, 30, 34

Appendix 4B

Constructing RIASEC Profiles for Third-Year Secondary Education Programs

To define the interest profiles of distinct study environments in the third year of secondary education, a sample-based approach was employed. Interest assessments were carried out over three consecutive academic years, beginning in 2021-2022, targeting third-year students to provide a comprehensive understanding of interest profiles within the target environment. The questionnaires were administered between January and May to ensure that students had completed at least one exam period. A total of 2,872 participants provided consent and completed the SIMON-jr interest inventory. For analysis, we included only the RIASEC scores of students who expressed a clear willingness to select their current study program again or a tentative willingness if they found the program interesting. This selection criterion resulted in a final sample of 2,486 participants (M_{age} = 14.43, SD_{age} = .67), of whom 46% were female (1% was unspecified), and 7% spoke a home language other than Dutch. Participants were drawn from various locations across Flanders. The RIASEC scores of students in each study environment were averaged across all six dimensions to generate a RIASEC profile for each environment. The SIMON-jr showed sufficient to good internal consistency across all RIASEC scales, with Cronbach alphas of .83, .88, .85, .83, .76, .67 for the respective interest dimensions. The number of students contributing to each study environment profile varied from 22 to 301, with a mean of 95.56 and standard deviation of 74.68.

Appendix 4C

Table 4C1

Standardized Factor Loadings (SE) for Six-factor CFA

Indicator	Factor 1 (R)	Factor 2 (I)	Factor 3 (A)	Factor 4 (S)	Factor 5 (E)	Factor 6 (C)
R1	.66					
R2	.79					
R3	.96					
R4	.77					
R5	.82					
R6	.81					
I1		.63				
I2		.86				
I3		.67				
I4		.91				
I5		.88				
I6		.92				
A1			.71			
A2			.94			
A3			.94			
A4			.93			
A5			.64			
A6			.68			
S1				.72		
S2				.79		
S3				.85		
S4				.75		
S5				.82		
S6				.78		
E1					.64	
E2					.55	
E3					.65	
E4					.70	
E5					.78	
E6					.83	
C1						.64
C3						.71
C3						.68
C4						.59
C5						.62
C6						.57

Appendix 4D

Table 4D1

Results of Differential Item Functioning Analysis

Item n°	Mantel-Haenszel χ^2	*p*	log odds ratio effect size (DeltaMH)
R1	2.20	.138	0.85 (A)
R2	2.85	.091	1.20 (B)
R3	2.97	.085	-1.28 (B)
R4	4.32	.038	1.22 (B)
R5	0.29	.587	-0.46 (A)
R6	14.22	<.001	-2.50 (C)
I1	17.85	<.001	-1.94 (C)
I2	5.16	.023	1.21 (B)
I3	2.93	.087	0.82 (A)
I4	0.00	.978	0.06 (A)
I5	0.13	.724	0.25 (A)
I6	0.00	.949	-0.13 (A)
A1	14.18	<.001	1.89 (C)
A2	0.19	.661	-0.36 (A)
A3	0.93	.334	-0.67 (A)
A4	1.61	.205	-0.99 (A)
A5	0.30	.582	0.31 (A)
A6	4.43	.035	-1.15 (B)
S1	0.28	.598	-0.42 (A)
S2	0.17	.681	-0.28 (A)
S3	0.55	.459	0.56 (A)
S4	11.08	<.001	-2.10 (C)
S5	0.32	.569	0.39 (A)
S6	7.45	.006	1.61 (C)
E1	3.63	.057	0.88 (A)
E2	3.63	.057	0.92 (A)
E3	1.28	.258	0.57 (A)
E4	4.33	.037	-1.00 (B)
E5	2.29	.130	-0.90 (A)
E6	3.80	.051	-1.08 (B)
C1	0.05	.831	-0.20 (A)
C2	1.39	.239	-0.68 (A)
C3	0.25	.616	-0.35 (A)
C4	3.40	.065	-1.12 (B)
C5	10.46	.001	1.70 (C)
C6	0.15	.697	0.23 (A)

Note. Effect size code A = negligible effect; B = moderate effect; C = large effect (Magis et al., 2010).

Appendices Chapter 5

Appendix 5A

Table 5A1

Demographic Characteristics of Students who Completed all Three Waves Versus Students who did not Complete all Three Waves

	Students who completed all three waves	Students who did not complete all three waves
% girls	48%	40%
% home language is not Dutch	7%	12%
% with scholarship	25%	31%
% highest attained degree is primary education or less Parent 1	9%	12%
% highest attained degree is primary education or less Parent 2	8%	12%

Appendix 5B

Analyses of Mean-level Changes in Vocational Interests

Table 5B1 presents descriptive statistics for the RIASEC dimensions among participants who completed all three measurement waves, including means and standard deviations for each wave, as well as gender differences.

Table 5B1

Descriptive Statistics of RIASEC Dimensions across Three Waves by Gender

	M Overall	*SD* Overall	*M* Boys	*SD* Boys	*M* Girls	*SD* Girls	*M* diff. (boys – girls)
Realistic T1	28.28	31.91	43.95	33.89	11.40	18.22	32.55
Realistic T2	25.14	32.87	39.25	35.28	9.94	21.53	29.31
Realistic T3	24.65	33.20	37.90	36.01	10.45	22.49	27.45
Investigative T1	43.20	35.64	47.58	34.99	38.23	35.73	9.35
Investigative T2	42.78	37.27	44.69	38.17	40.86	36.36	3.83
Investigative T3	40.17	39.06	39.92	38.84	40.79	39.37	-0.87
Artistic T1	36.58	34.12	20.16	24.39	54.17	34.43	-34.01
Artistic T2	28.63	33.64	14.72	23.51	43.64	36.48	-28.92
Artistic T3	26.19	32.85	13.84	22.87	39.55	36.74	-25.71
Social T1	77.30	29.48	67.14	32.11	88.30	21.61	-21.16
Social T2	68.62	33.14	55.85	33.64	82.46	26.53	-26.61
Social T3	64.12	35.27	47.31	34.68	82.31	25.73	-35.00
Enterprising T1	40.41	30.24	48.12	29.02	32.09	29.47	16.03
Enterprising T2	43.41	33.48	52.22	32.70	33.92	31.71	18.3
Enterprising T3	46.62	34.17	54.50	33.75	38.30	32.66	16.2
Conventional T1	19.70	23.51	22.31	23.51	16.96	23.31	5.35
Conventional T2	21.06	25.46	22.72	26.60	19.37	24.16	3.35
Conventional T3	24.72	27.12	25.34	27.49	24.12	26.83	1.22

Note. $N_{overall}$ = 478; N_{boys} = 248; N_{girls} = 228.

The ICCs revealed that a substantial proportion of the variance in vocational interests is attributable to differences between students: 67%, 62%, 63%, 51%, 53%, and 48%, in the order of the RIASEC-acronym. These results confirm that multilevel models are highly suitable for analyzing vocational interest development patterns.

Table 5B2 presents the results of the multilevel models for all RIASEC scales. Pairwise contrast results for all RIASEC dimensions are provided in Tables 5B3 to 5B8. Additionally, Figure 5B1 visualizes the mean-level interest trajectories for each RIASEC dimension over time, separated by gender.

The conditional random-intercept models, which included time, gender and the interaction between time and gender, revealed several significant findings across the RIASEC dimensions. Gender had a significant main effect on all RIASEC scales, as did time (wave), except for the Conventional dimension. Moreover, significant interaction effects between time and gender were observed for all dimensions except for Enterprising, suggesting that the rate of change over time differed between genders for most RIASEC dimensions.

Pairwise contrasts were conducted to explore these patterns further. While the conditional model indicated a small but significant main effect of gender in the Conventional dimension, follow-up contrasts did not detect significant differences between gender groups at specific time points. This discrepancy may stem from limited statistical power in the contrasts or from the correction for multiple comparisons, which applies a stricter significance threshold. In the Investigative dimension, the gender difference was significant at T1 but disappeared in subsequent waves. The largest gender differences were observed in the Realistic, Artistic, and Social dimensions, with the Social dimension being the only dimension where the gender gap widened over time.

Contrasts examining mean-level changes within each gender revealed a general downward trend in the Realistic, Investigative, Artistic, and Social dimensions, with an upward trend in the Enterprising dimension. However, the nature of these changes differed by gender and time interval. The declines in Realistic and Investigative interests were significant for boys, but not for girls. Boys exhibited significant declines in Social interest across all time points, while for girls, the decline was only significant between T1 and T2. Only girls showed a significant increase in Conventional interest. The largest decreases were observed in Social interest for boys and in Artistic interest for girls.

Table 5B2

Results of the Multilevel Models for all RIASEC Scales

	Realistic Interest				Investigative Interest				Artistic Interest			
	β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*
Fixed Effects												
Intercept	43.12	1.35	31.85	<.001	44.88	1.70	26.45	<.001	20.27	1.41	14.38	<.001
Wave1	-5.17	1.35	-3.84	<.001	-3.81	1.62	-2.35	.019	-6.12	1.46	-4.18	<.001
Wave2	-6.94	1.48	-4.69	<.001	-7.99	1.78	-4.48	<.001	-6.28	1.61	-3.90	<.001
Gender1	-29.98	2.01	-14.93	<.001	-8.11	2.52	-3.22	.001	34.55	2.09	16.53	<.001
Wave1:Gender1	1.74	1.97	.89	.38	4.34	2.37	1.83	.067	-4.55	2.14	-2.12	.03
Wave2:Gender1	4.77	2.17	2.20	.03	10.35	2.61	3.97	<.001	-9.09	2.36	-3.86	<.001
Random Effects	Variance	SD			Variance	SD			Variance	SD		
ID (Intercept)	503.00	22.43			836.50	28.92			505.80	22.49		
Residual	342.80	18.51			495.40	22.26			408.50	20.21		
Marginal R^2	.19				.01				.22			
Conditional R^2	.67				.63				.65			

	Social Interest				Enterprising Interest				Conventional Interest			
	β	SE	*t*	*p*	β	SE	*t*	*p*	β	SE	*t*	*p*
Fixed Effects												
Intercept	66.86	1.42	47.03	<.001	47.82	1.45	32.98	<.001	21.88	1.17	18.64	<.001
Wave1	-13.66	1.60	-8.56	<.001	2.39	1.57	1.52	.13	-.41	1.30	-.31	.754
Wave2	-19.82	1.75	-11.31	<.001	4.65	1.72	2.70	.01	2.24	1.42	1.58	.115
Gender1	18.31	2.11	8.69	<.001	-15.05	2.15	-7.00	<.001	-4.34	1.74	-2.50	.013
Wave1:Gender1	8.14	2.34	3.48	<.001	-1.53	2.30	-.67	.51	1.58	1.90	.83	.405
Wave2:Gender1	14.49	2.57	5.65	<.001	1.37	2.53	.54	.59	4.15	2.08	1.99	.047
Random Effects	Variance	SD			Variance	SD			Variance	SD		
ID (Intercept)	434.50	20.84			493.40	22.21			309.30	17.59		
Residual	490.50	22.15			471.70	21.72			322.50	17.96		
Marginal R^2	.17				.06				.01			
Conditional R^2	.56				.54				.49			

Note. The fixed effects intercept represents the baseline value of the respective interest dimension for boys at T1

Table 5B3

Results of Pairwise Contrasts for Realistic Interests Across Waves and Gender

Contrast		Estimate	SE	*p*
T1	boys - girls	29.98	2.01	<.001
T2	boys - girls	28.23	2.12	<.001
T3	boys - girls	25.20	2.30	<.001
Boys	T1 - T2	5.17	1.34	.002
	T2 -T3	1.77	1.50	1.00
	T1 - T3	6.94	1.48	<.001
Girls	T1 - T2	3.42	1.44	.26
	T2 -T3	-1.26	1.61	1.00
	T1 - T3	2.17	1.58	1.00

Table 5B4

Results of Pairwise Contrasts for Investigative Interests Across Waves and Gender

Contrast		Estimate	SE	*p*
T1	boys - girls	8.11	2.52	.02
T2	boys - girls	3.77	2.65	1.00
T3	boys - girls	-2.24	2.86	1.00
Boys	T1 - T2	3.81	1.62	.28
	T2 -T3	4.18	1.81	.31
	T1 - T3	7.99	1.78	<.001
Girls	T1 - T2	-.54	1.73	1.00
	T2 -T3	-1.83	1.93	1.00
	T1 - T3	-2.36	1.90	1.00

Table 5B5

Results of Pairwise Contrasts for Artistic Interests Across Waves and Gender

Contrast		Estimate	SE	*p*
T1	boys - girls	-34.55	2.09	<.001
T2	boys - girls	-30.00	2.22	<.001
T3	boys - girls	-25.46	2.42	<.001
Boys	T1 - T2	6.12	1.46	<.001
	T2 -T3	.16	1.63	1.00
	T1 - T3	6.28	1.61	.002
Girls	T1 - T2	10.67	1.57	<.001
	T2 -T3	4.70	1.75	.111
	T1 - T3	15.37	1.72	<.001

Table 5B6

Results of Pairwise Contrasts for Social Interests Across Waves and Gender

Contrast		Estimate	SE	*p*
T1	boys - girls	-18.31	2.11	<.001
T2	boys - girls	-26.45	2.25	<.001
T3	boys - girls	-32.81	2.48	<.001
Boys	T1 - T2	13.66	1.60	<.001
	T2 -T3	6.17	1.78	.009
	T1 - T3	19.82	1.75	<.001
Girls	T1 - T2	5.52	1.71	.019
	T2 -T3	-.19	1.91	1.00
	T1 - T3	5.33	1.88	.068

Table 5B7

Results of Pairwise Contrasts for Enterprising Interests Across Waves and Gender

Contrast		Estimate	SE	*p*
T1	boys - girls	15.05	2.15	<.001
T2	boys - girls	16.58	2.29	<.001
T3	boys - girls	13.68	2.51	<.001
Boys	T1 - T2	-2.39	1.57	1.00
	T2 -T3	-2.27	1.75	1.00
	T1 - T3	-4.65	1.72	.11
Girls	T1 - T2	-.86	1.68	1.00
	T2 -T3	-5.17	1.88	.09
	T1 - T3	-6.02	1.84	.02

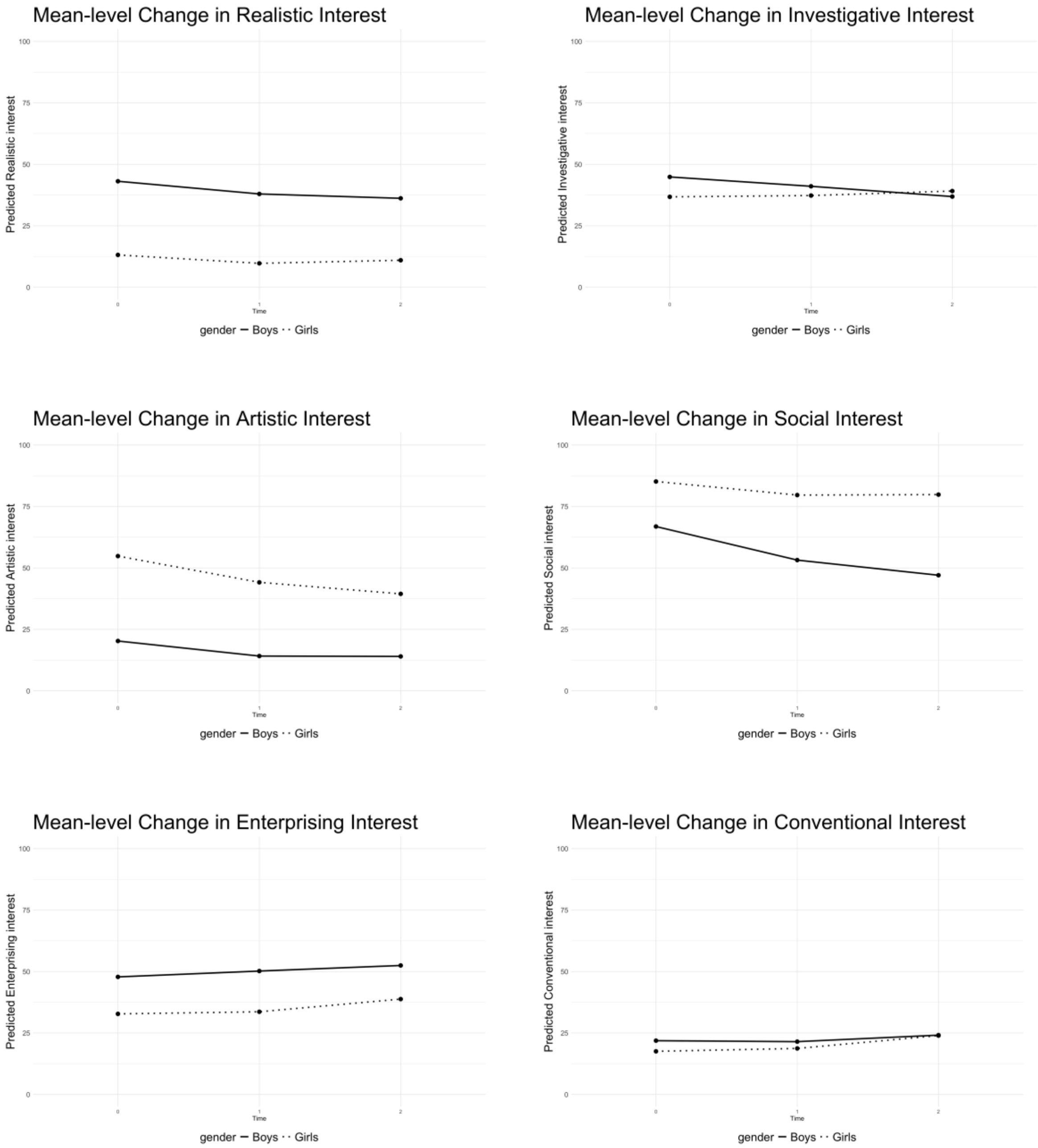
Table 5B8

Results of Pairwise Contrasts for Conventional Interests Across Waves and Gender

Contrast		Estimate	SE	*p*
T1	boys - girls	4.34	1.74	.190
T2	boys - girls	2.76	1.86	1.00
T3	boys - girls	.19	2.04	1.00
Boys	T1 - T2	.41	1.30	1.00
	T2 -T3	-2.65	1.45	1.00
	T1 - T3	-2.24	1.42	1.00
Girls	T1 - T2	-1.18	1.39	1.00
	T2 -T3	-5.22	1.55	.012
	T1 - T3	-6.40	1.52	<.001

Figure 5B1

Mean-level Changes per Gender across all RIASEC Dimensions



Data Storage Fact Sheets

Chapter 2

Name/identifier study: Doctoral dissertation: Testing talent: validating tools for informed study choices in secondary education – Chapter 2

Author: Merel Dutry

Date: February 5, 2025

1. Contact details

==

1a. Main researcher

- name: Merel Dutry
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: merel.dutry@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Wouter Duyck
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==

* Reference of the publication in which the datasets are reported:

Dutry, M., Vereeck, A., Duyck, W., Derous, E., Schelfhout, S., Szmalec, A., Woumans, E., Schittekatte, M., Debeer, D., & Dirix, N. (2025). Validation of the Children's International Cognitive Ability Resource (Ch-ICAR). *Behavior Research Methods, 57.* https://doi.org/10.3758/s13428-024-02591-1

* Which datasets in that publication does this sheet apply to?:

All data

3. Information about the files that have been stored

==

3a. Raw data

----

* Have the raw data been stored by the main researcher? [x] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

  - [x] researcher PC

  - [x] research group file server

  - [x] other (specify): external hard drive

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [x] main researcher

  - [x] responsible ZAP

- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify):

3b. Other files

----

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to reported results. Specify: ...
  - [x] file(s) containing processed data. Specify: Excel files
  - [x] file(s) containing analyses. Specify: R codes available at OSF via:

https://osf.io/6cwfs/?view_only=4a80feb83bd94c378d3bd06853820563
  - [x] files(s) containing information about informed consent
  - [x] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files and how this content should be

interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [x] individual PC
  - [x] research group file server
  - [x] other: external hard drive

* Who has direct access to these other files (i.e., without intervention of another person)?
  - [x] main researcher
  - [x] responsible ZAP

- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==

* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):

  - name:
  - address:
  - affiliation:
  - e-mail:

Chapter 3

Name/identifier study: Doctoral dissertation: Testing talent: validating tools for informed study choices in secondary education – Chapter 3

Author: Merel Dutry

Date: February 12, 2025

1. Contact details

===

1a. Main researcher

- name: Merel Dutry

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: merel.dutry@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Wouter Duyck

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==

* Reference of the publication in which the datasets are reported:

Merel Dutry, Stijn Schelfhout, Eva Derous, Wouter Duyck, Klaas Bombeke, Jonas De Bruyne, Aleksandra Zheleva, Jelle Demanet, and Nicolas Dirix: Validating CalibrainVR-MR: Applying a VR-based mental rotation test to study orientation in secondary education. Submitted to *Psychologica Belgica*, April 2025

* Which datasets in that publication does this sheet apply to?:

All data

3. Information about the files that have been stored

==

3a. Raw data

----

* Have the raw data been stored by the main researcher? [x] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

  - [x] researcher PC

  - [x] research group file server

  - [x] other (specify): external hard drive

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [x] main researcher

  - [x] responsible ZAP

  - [ ] all members of the research group

  - [ ] all members of UGent

  - [ ] other (specify):

3b. Other files

----

* Which other files have been stored?

  - [ ] file(s) describing the transition from raw data to reported results. Specify: ...

  - [x] file(s) containing processed data. Specify: Excel files

  - [x] file(s) containing analyses. Specify: R codes available at OSF via:

https://osf.io/4evwt/?view_only=ff58b13d833f435aa6ab6b4f551708a4

  - [x] files(s) containing information about informed consent

  - [x] a file specifying legal and ethical provisions

  - [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...

  - [ ] other files. Specify: ...

* On which platform are these other files stored?

  - [x] individual PC

  - [x] research group file server

  - [x] other: external hard drive

* Who has direct access to these other files (i.e., without intervention of another person)?

  - [x] main researcher

  - [x] responsible ZAP

- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

===

* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail:

Chapter 4

Name/identifier study: Doctoral dissertation: Testing talent: validating tools for informed study choices in secondary education – Chapter 4

Author: Merel Dutry

Date: February 12, 2025

1. Contact details

===

1a. Main researcher

- name: Merel Dutry

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: merel.dutry@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Wouter Duyck

- address: Henri Dunantlaan 2, 9000 Gent

- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

==

* Reference of the publication in which the datasets are reported:

Merel Dutry, Stijn Schelfhout, Eva Derous, Wouter Duyck, and Nicolas Dirix: Optimizing Career Guidance in Secondary Education: Exploring Person-Environment Interest Fit with the SIMON-jr Interest Inventory.

Submitted to *Journal of Career Assessment,* April 2025

* Which datasets in that publication does this sheet apply to?:

All data

3. Information about the files that have been stored

==

3a. Raw data

----

* Have the raw data been stored by the main researcher? [x] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

  - [x] researcher PC
  - [x] research group file server
  - [x] other (specify): external hard drive

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [x] main researcher
  - [x] responsible ZAP

- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify):

3b. Other files

----

* Which other files have been stored?

- [ ] file(s) describing the transition from raw data to reported results. Specify: ...
- [x] file(s) containing processed data. Specify: Excel files
- [x] file(s) containing analyses. Specify: R codes available at OSF via: https://osf.io/esf8y/?view_only=aae3a86696c543c380b01dd3b375201d.
- [x] files(s) containing information about informed consent
- [x] a file specifying legal and ethical provisions
- [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
- [ ] other files. Specify: ...

* On which platform are these other files stored?

- [x] individual PC
- [x] research group file server
- [x] other: external hard drive

* Who has direct access to these other files (i.e., without intervention of another person)?

- [x] main researcher
- [x] responsible ZAP

- [ ] all members of the research group
- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==

* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
   - name:
   - address:
   - affiliation:
   - e-mail:

Chapter 5

Name/identifier study: Doctoral dissertation: Testing talent: validating tools for informed study choices in secondary education – Chapter 5

Author: Merel Dutry

Date: February 12, 2025

1. Contact details

===

1a. Main researcher

- name: Merel Dutry
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: merel.dutry@ugent.be

1b. Responsible Staff Member (ZAP)

- name: Wouter Duyck
- address: Henri Dunantlaan 2, 9000 Gent
- e-mail: wouter.duyck@ugent.be

If a response is not received when using the above contact details, please send an email to data.pp@ugent.be or contact Data Management, Faculty of Psychology and Educational Sciences, Henri Dunantlaan 2, 9000 Ghent, Belgium.

2. Information about the datasets to which this sheet applies

===

* Reference of the publication in which the datasets are reported:

Merel Dutry, Stijn Schelfhout, Eva Derous, Wouter Duyck, and Nicolas Dirix: Interest and Person-Environment Interest Fit Development in Early Adolescence.

Submitted to *Journal of Vocational Behavior*, April 2025

* Which datasets in that publication does this sheet apply to?:

All data

3. Information about the files that have been stored

===

3a. Raw data

----

* Have the raw data been stored by the main researcher? [x] YES / [ ] NO

If NO, please justify:

* On which platform are the raw data stored?

  - [x] researcher PC

  - [x] research group file server

  - [x] other (specify): external hard drive

* Who has direct access to the raw data (i.e., without intervention of another person)?

  - [x] main researcher

  - [x] responsible ZAP

  - [ ] all members of the research group

- [ ] all members of UGent
- [ ] other (specify):

3b. Other files

----

* Which other files have been stored?
  - [ ] file(s) describing the transition from raw data to reported results. Specify: ...
  - [x] file(s) containing processed data. Specify: Excel files
  - [x] file(s) containing analyses. Specify: R codes available at OSF via: https://osf.io/2fvy8/?view_only=78eb66cabd544088aba1b7528b072cf3.
  - [x] files(s) containing information about informed consent
  - [x] a file specifying legal and ethical provisions
  - [ ] file(s) that describe the content of the stored files and how this content should be interpreted. Specify: ...
  - [ ] other files. Specify: ...

* On which platform are these other files stored?
  - [x] individual PC
  - [x] research group file server
  - [x] other: external hard drive

* Who has direct access to these other files (i.e., without intervention of another person)?
  - [x] main researcher
  - [x] responsible ZAP
  - [ ] all members of the research group

- [ ] all members of UGent
- [ ] other (specify): ...

4. Reproduction

==

* Have the results been reproduced independently?: [ ] YES / [x] NO

* If yes, by whom (add if multiple):
  - name:
  - address:
  - affiliation:
  - e-mail: